Quantifying Texture Scale in Accordance With Human Perception

Sami Bouremoum

A dissertation submitted in partial fulfilment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

UCL CoMPLEX - Department of Computer Science

University College London

May, 2018

I, Sami Bouremoum, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Date: Wednesday 9th May, 2018

Signed:

Abstract

Visual texture has multiple perceptual attributes (e.g. regularity, isotropy, etc.), including scale. The scale of visual texture has been defined as the size of the repeating unit (or texel) of which the texture is composed. Not all textures are formed through the placement of a clearly discernible repeating unit (e.g. irregular and stochastic textures). There is currently no rigorous definition for texture scale that is applicable to textures of a wide range of regularities. We hypothesised that texture scale ought to extend to these less regular textures. Non-overlapping sample windows (or patches) taken from a texture appear increasingly similar as the size of the window gets larger. This is true irrespective of whether the texture is formed by the placement of a discernible repeating unit or not. We propose the following new characterisation for texture scale: "the smallest window size beyond within which texture appears consistently". We perform two psychophysical studies and report data that demonstrates consensus across subjects and across methods of probing in the assessment of texture scale. We then present an empirical algorithm for the estimation of scale based on this characterisation. We demonstrate agreement between the algorithm and (subjective) human assessment with an RMS accuracy of 1.2 just-noticeable-differences, a significant improvement over previous published algorithms. We provide two ground-truth perceptual datasets, one for each of our psychophysical studies, for the texture scale of the entire Brodatz album, together with confidence levels for each of our estimates. Finally, we make available an online tool which researchers can use to obtain texture scale estimates by uploading images of textures.

Acknowledgements

I would like to express my gratitude towards my supervisors, Lewis D. Griffin and Buzz Baum, for their guidance and support throughout my PhD. They have always been enthusiastic, supportive and inspiring, far beyond what is asked of them. This thesis would not have been possible without Lewis's encouragement in overcoming my motivational lulls.

I have also received immense support from both of UCL's academic and administrative staff and, in particular, from CoMPLEX and the department of Computer Science. During my PhD I have had the opportunity to immerse myself in the wider UCL community through which I made many great friends and built a fantastic network of interesting people.

I would like to thank EPSRC for funding this research through UCL CoMPLEX,

I would also like to thank my family and friends who so willingly offered their support and volunteered in our perceptual studies - it is not fascinating to click through hundreds of images and make judgments on their scale, but they did so with utmost diligence and dedication. I would especially like to thank my wife, Sophie, who is always there for me with words of encouragement when times are toughest and also my parents who regularly reminded me of the importance of seeing my PhD through to the end. The PhD journey is one on which I will look back in the future years and for which I will be (and am) incredibly grateful.

Contents

1 Introduction **13**

1.1 Motivation 16

1.2 Contribution 18

1.3 Thesis Outline 19

1.4 Publications and Dissemination 20

2 Background Material **22**

2.1 Texture, Scale, Regularity and Periodicity 22

2.2 Measurement Theory 29

2.3 Psychophysics 30

2.4 Textures, Databases and Applications 35

2.4.1 Natural Images 36

2.4.2 Material Textures 39

2.4.3 Texture in the Biosciences 43

2.4.4 Dynamic Textures 45

2.5 A Taxonomy of Texture Representations 47

2.5.1 Statistical Texture Methods 47

2.5.2 Transform Based Methods 50

2.5.3 Model-Based Methods 51

2.5.4 Structural/Geometric Methods 52

2.5.5 Machine Learning Methods 53

2.6 Relevant Methods for Texture Scale 53

2.7 Summary from Literature Review 58

3 Quantifying Human Perception of Texture Scale 60

3.1 Choosing a Suitable Dataset . . . 63

3.2 Direct Scaling Approach . . . 64

3.2.1 Method . . . 64

3.2.2 Response Analysis . . . 66

3.2.3 Summary . . . 71

3.3 Indirect Scaling Approach . . . 72

3.3.1 Method . . . 74

3.3.2 Building an Interval Scale . . . 75

3.3.3 Response Analysis . . . 78

3.3.4 Summary . . . 79

3.4 Comparison of the Two Approaches . . . 81

3.5 Summary of Perceptual Studies . . . 83

4 Estimating Texture Scale 85

4.1 Calculating Texture Curves . . . 86

4.1.1 Representations . . . 87

4.1.1.1 Histograms of Gray-Scale Values . . . 88

4.1.1.2 Keypoint Maps - SURF . . . 89

4.1.1.3 Basic Image Features . . . 90

4.1.2 Texture Curves . . . 95

4.1.2.1 The Odd-One-Out Texture Curves . . . 97

4.1.2.2 The Mean Histogram Distance Texture Curves . . . 99

4.1.2.3 The Total Histogram Variance Texture Curves . . . 100

4.1.2.4 SURF Keypoint Curves . . . 104

4.1.2.5 Discussion on Texture Curves . . . 104

4.2 Algorithm for Estimating Texture Scale . . . 106

4.2.1 Choice of Curve and Data . . . 106

4.2.2 Translating Curve Data to Texture Scale Estimates . . . 108

4.2.3 Performance Assessment . . . 109

4.3 Algorithm for Estimating Texture Scale . . . 110

4.3.1 Complete Method 111
4.3.2 Simplified Method 112
4.4 Results and Performance 113
4.5 Summary and Discussion 117

5 An Online Tool for Calculating Texture Scale 121
5.1 The Interface 121
5.2 User Feedback 123
5.2.1 First User Test 123
5.2.2 Second User Test 126
5.3 Summary 128

6 Summary & Conclusions 130
6.1 Critical Appraisal of Contribution 132
6.1.1 A New Characterisation of Texture Scale 133
6.1.2 Evidence of Human Ability to Assess Texture Scale Irrespective of Regularity 133
6.1.3 Perceptually Validated Texture Scale Values for the Brodatz 134
6.1.4 Algorithm for Estimating Texture Scale 135
6.1.5 Online Texture Scale Calculator 135
6.2 Suggested Direction of Further Research 136

Appendices 138

Bibliography 139

List of Figures

1.1 Six examples of visual texture 13
1.2 Increasing size samples from a texture look increasingly consistent . 15
1.3 Three distinct textured regions on Drosophila eye 17

2.1 Taxonomy for texture regularity 25
2.2 Twelve original Brodatz textures 36
2.3 Four examples from the coloured Brodatz dataset 37
2.4 Two USC-SIPI texture tiles and corresponding segmentations 38
2.5 Examples from the CUReT database 40
2.6 Examples from the KTH-Tips database 41
2.7 Examples from the PhoTex database 42
2.8 Examples from the IICBU Biological Image Repository 44
2.9 Four retina images from MESSIDOR 45
2.10 Keypoint locations for two Brodatz textures 54
2.11 Keypoint density maps by Ardizzone et al. (2013) 55
2.12 Number of keypoints *vs.* window size. 55
2.13 Tiling method by Hong et al. (2008) 57
2.14 Texture scale by Hong et al. (2008) 58
2.15 Texture scale map by Hong et al. (2008) 59

3.1 Direct-scaling experiment interface 66
3.2 Histogram of subjects' responses to direct-scaling experiment . . . 67
3.3 Systematic subject differences seen in subject-pair scatter plots . . . 67
3.4 Sample mean and CI responses to direct-scaling experiment 69

3.5 Standard deviation of responses to direct-scaling experiment 69
3.6 Four Brodatz with largest associated response variance 70
3.7 Distribution of time taken per response in direct-scaling experiment 70
3.8 Location of subject selections in direct-scaling experiment 70
3.9 Variation of responses to direct-scaling experiment by isotropy . . . 71
3.10 Lower variance in subject's responses for isotropic textures 72
3.11 Indirect-scaling experiment interface 75
3.12 Four Brodatz textures on an interval scale 76
3.13 Four Brodatz with largest response variance in 2AFC experiment . . 77
3.14 Histogram of deviances from goodness-of-fit analysis 78
3.15 Texture scale and CI for Brodatz textures from 2AFC experiment . . 79
3.16 Variation of responses to 2AFC task according to isotropy 81
3.17 Comparison of scale estimates and SDs between two experiments . 82
3.18 Comparison of scales between experiments according to isotropy . . 83

4.1 Illustration of the concept of texture curves 87
4.2 Histogram of gray-scale values 88
4.3 SURF keypoints for Brodatz D83 89
4.4 Variation of number of keypoints with window size 90
4.5 Gaussian blur of Brodatz D83 91
4.6 Second order structure solid - BIFs 93
4.7 BIF variation with σ_B and γ 95
4.8 Illustrative BIF histogram 96
4.9 Larger samples from texture have more similar BIF histograms . . . 97
4.10 Odd-One-Out calculation - visual illustration 97
4.11 OOO curve for Brodatz D83 98
4.12 OOO curves for single scale *vs.* column BIFs 99
4.13 Mean Histogram Distance calculation - visual illustration 100
4.14 Mean Histogram Distance curves for all Brodatz and 25 BIF settings 101
4.15 Mean Histogram Distance curves for Brodatz D7 and D8 101
4.16 Total Histogram Variance curves for all Brodatz and 25 BIF settings 102

4.17 THV curves for Brodatz D38 and D102 103
4.18 Mean Log THV curves for Brodatz collection 103
4.19 SURF keypoint density curves 105
4.20 Candidate metrics from texture curves 107
4.21 Candidate maps from curve data to perceptual ground-truth 109
4.22 Scatter plot of algorithm derived scale *vs.* 2AFC ground-truth 114
4.23 Algorithm estimates *vs.* 2AFC coloured according to regularity . . . 114
4.24 Algorithm estimates *vs.* 2AFC coloured according to isotropy . . . 115
4.25 Texels for 24 Brodatz formed by a discernible repeating element . . 119

5.1 Interface landing page 122
5.2 Interface output 124
5.3 Feedback form 125
5.4 Histogram of feedback responses 126
5.5 Three textures for which we received poor performance feedback . . 127
5.6 Histogram of feedback for A-B testing 128

List of Tables

3.1 Brodatz textures scored according to regularity and isotropy 65

3.2 Brodatz $\log_2$ texture scales and SDs from direct-scaling approach . . 73

3.3 Brodatz texture scales and SDs from direct-scaling approach 74

3.4 2AFC texture scales and CI for Brodatz in jnd units 80

3.5 2AFC derived scale estimates and CIs for Brodatz in $\log_2$ pixels . . 84

4.1 Algorithm performance comparison 116

4.2 Perceptual scale ground-truth values in jnds and pixels 118

4.3 Correlation between texture scale estimates - algorithms and perceptual. 120

Abbreviations

2AFC . . . . . 2 Alternative Forced Choice

jnd . . . . . Just noticeable difference

DoG . . . . . Derivative of Gaussian

RMS . . . . . Root Mean Square

RMSE . . . . . Root Mean Square Error

BIF . . . . . Basic Image Features

OOO . . . . . Odd One Out

MHD . . . . . Mean Histogram Distance

THV . . . . . Total Histogram Variance

SD . . . . . Standard Deviation

CI . . . . . Confidence Interval

GoF . . . . . Goodness of Fit

ML . . . . . Maximum Likelihood

Chapter 1

Introduction

We are all familiar with the concept of texture in the tactile / haptic sense. The moist sponginess of a cake, the softness of velvet, and the roughness of a pineapple's skin are all examples of texture. As light reflects off these surfaces, the variation in light intensity reaching our eyes enables us to perceive texture visually. Such *visual texture* can also be perceived from flat surfaces, such as wood grain, or painted patterns. Henceforth, when we say *texture*, we will be referring to *visual texture*. Figure 1.1 shows some examples of visual textures found in the real world, and in the lab.

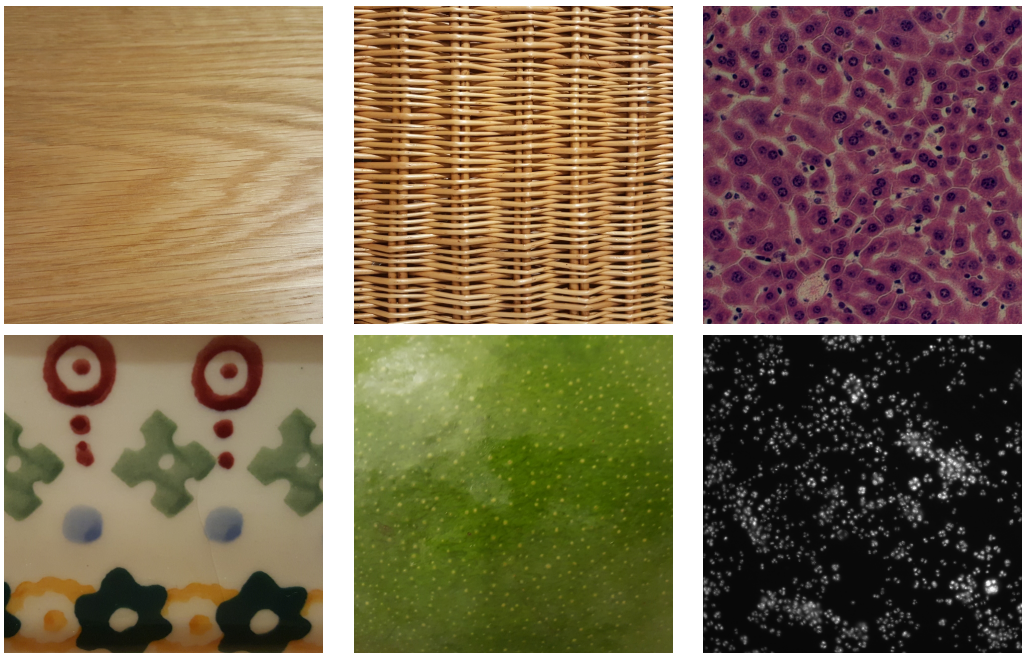


Figure 1.1: Six images of textured or patterned surfaces.

Texture has multiple attributes, such as regularity, isotropy, roughness, coarseness and scale. Some attributes have had algorithms developed for their quantification, and some have been validated psychophysically. We are interested in texture scale. Scale, when referring to texture has been defined in numerous ways. It can be defined as the size of a repeating unit, or equivalently, as the mean distance between the repeating units (Malik et al., 2001). Ardizzone et al. (2013) define the scale of texture as the smallest upright square window containing a consistent number of image keypoints. Hong et al. (2008) define scale as the smallest upright square window for which tiled neighbouring windows of the same size are similar.

All of these definitions have their limitations. A limitation of the definition by Malik et al. (2001) is that it requires the texture to be formed of a repeating unit placed in a grid-like fashion on the frame. Scale is thus ill-defined for non-regular textures. The definitions given by Ardizzone et al. (2013) and Hong et al. (2008) aim to identify the smallest upright square which could *a priori* apply to textures not formed of a repeating element placed on a grid, but *(i)* require the windows to be upright squares, and *(ii)* have not been validated perceptually.

Consider Figure 1.2. It shows upright square samples from the Brodatz texture D23 (Brodatz, 1966b) of increasing size. We have chosen upright square windows for illustration purposes, but these could be of any shape. It is clear that as the window size increases, the samples appear increasingly similar. To illustrate this further, consider the smallest window in the top half of Figure 1.2, and compare it to the smallest window in the bottom half of the figure. The two samples do not look consistent. Now consider the second smallest window in the top half of Figure 1.2 and compare it to the second smallest window in the bottom half of the figure. The two samples look more similar than the two smallest ones did. In fact, as we move from left to right on Figure 1.2, the samples appear increasingly consistent. Our hypothesis is therefore that there must be a smallest window size beyond within which texture appears consistently.

This hypothesis leads us to ask ourselves the following questions:

1. How do we characterise texture scale in such a way that it:

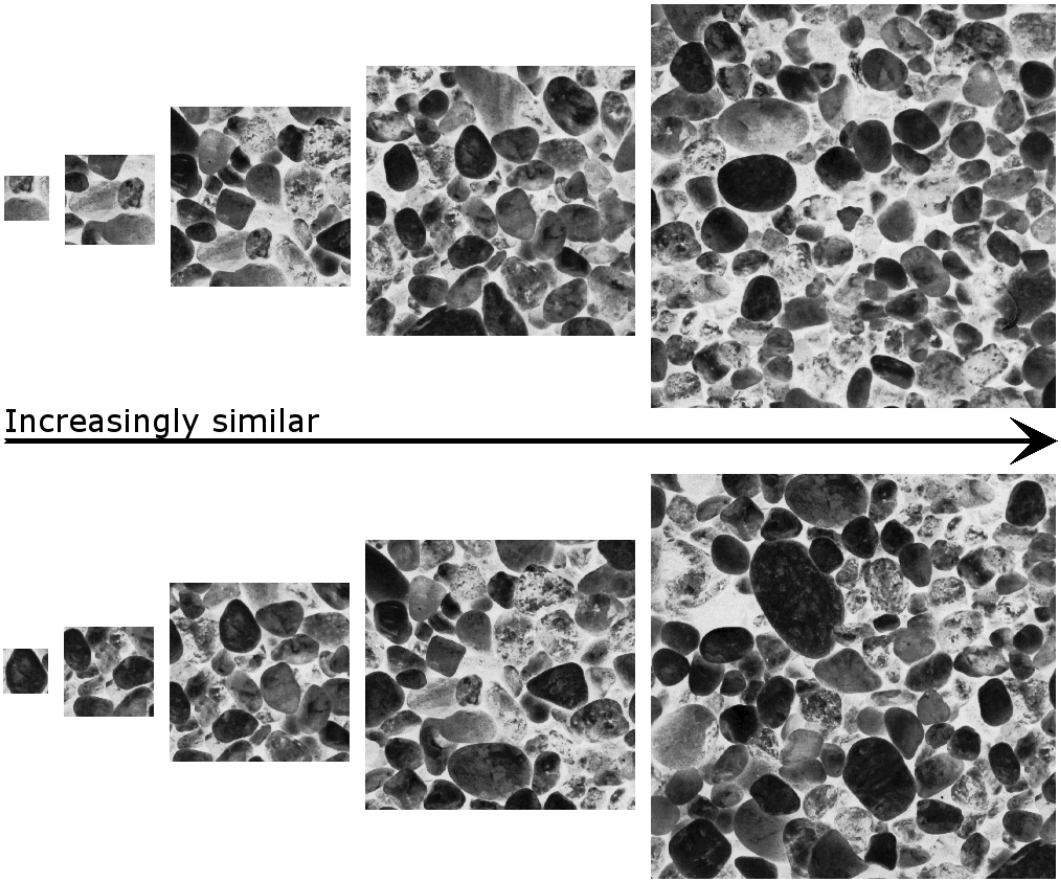


Figure 1.2: Increasing size samples from a texture look increasingly consistent.

(a) works for all textures;

(b) does not make any assumptions as to the structure of texture;

(c) can be used to verify our hypothesis of increasing similarity.

2. Provided we are able to characterise texture in such a way, are human observers able to assess the scale of textures consistently?

3. If so, what is a good way to capture human assessment of texture scale?

4. If we succeed in capturing it, can we devise an algorithm for estimating texture scale in a way that is consistent with human assessment?

5. If so, will researchers and other users find such a measure useful?

6. If they do, how can we enable experts from a range of disciplines, to benefit from our work?

To the best of our knowledge, there is no perceptually validated characterisation of texture scale in the literature that works for textures of all regularities. We hypothesise that there must be a smallest window size beyond within which texture appears consistently. We characterise texture scale as per our hypothesis: *the smallest window size beyond within which texture appears consistently*.

1.1 Motivation

Now that we have established our hypothesis, outlined the key questions that we need to answer, and defined the key properties we want our characterisation and algorithm to have, two further questions remain to be answered. The first is: "why should we care about the scale of non-regular textures?", and the second is: "why should we care whether estimates are perceptually validated or not?". We will answer the first question by providing two possible use-cases for texture scale: one hypothetical real-life example, and a second life-sciences based example. After introducing the two use-cases, we will explain why, for each of them, it is important for the derived scales to be perceptually validated.

Consider a merchant who trades in high-value luxury fabrics used for curtains and upholstering. Given the high-value of her merchandise, prospective customers come from far afield and require physical samples to be posted to them. The merchant has to establish what size samples she needs to send to her prospective customers. If the samples are too small, the customers will not be able to get a good idea of what a curtain or sofa made using her fabric would look like, and they are therefore less likely to make a purchase. If the samples are too large, she would be wasting money on the 90% of customers who will not purchase from her. The merchant needs to identify the scale of the fabrics so that she can send sample sizes that are large enough to convert enquiries into sales but not so large that sending samples becomes uneconomical.

Consider the image shown in Figure 1.3. It shows the cells of a new-born fruit-fly undergoing specialisation to form its cluster-eye. When fly larvae are born, they are completely blind. Where their eyes will later develop are undifferentiated cells

(Greenwood and Struhl, 1999), which look like the ones in the bottom-left part of Figure 1.3. Some time after birth, the cells develop a morphogenetic furrow which is a wave-like invagination of the cells caused by the constriction of the upper part of cells on a band. This wave is responsible for the specialisation of the undifferentiated cells, giving rise to the clusters that form the eye of the adult fly, as shown in the top-right of Figure 1.3 (Corrigall et al., 2007, Pichaud, 2014, Robertson et al., 2012).

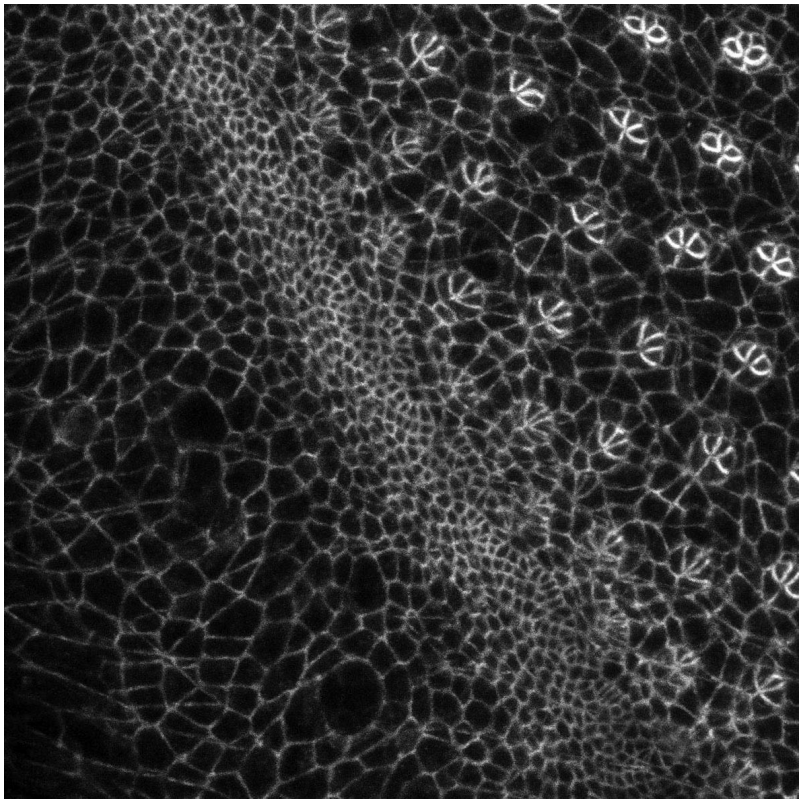


Figure 1.3: Drosophila eye undergoing morphogenetic furrow shows three distinct textured regions: Large scale disorganised cells in the bottom left of the image, compact disorganised cells on a diagonal band, and structured cells on the top right portion of the image. Courtesy of F. Pichaud Lab.

While the above description is concise, it is purely qualitative. Figure 1.3 can be treated as a texture. As we mentioned, texture has multiple attributes, such as regularity, isotropy, roughness, coarseness and scale. Protonotarios et al. (2014) propose an algorithm for the quantification of point patterns, which is validated against human perception. They use their approach to quantify the order of a cell tissue by treating each cell nucleus as a point. The three regions in Figure 1.3

show variation in scale as well as in order. For a more comprehensive quantitative description of the specialisation process, we would need to estimate the scale of each of the three different regions in Figure 1.3. Unfortunately, none of the existing definitions or methods for estimating texture scale are applicable to such images.

We have provided two use-cases for a measure of texture scale that are applicable to textures of varying regularities so we can move on to answering the second question: "why do we want our estimates to be perceptually validated?". In order for a measure to be useful to a wide range of disciplines, it needs to be communicable. What we mean by this is that the characterisation of scale ought to be understandable to the naïve reader. The definition proposed by Ardizzone et al. (2013), "The smallest upright square for which the distribution of the number of key-points becomes linear with the size of the square", may be difficult to interpret by an audience beyond the image analysis / vision spheres. We propose a characteristic that is perceptual, and so it needs to be validated perceptually to ensure that the estimates which our algorithm produces are meaningful to a reader.

1.2 Contribution

The five primary contributions presented in this thesis are the following:

We propose a new characterisation for texture scale that is devoid of any assumption with regards to the structure or definition of texture, and that is applicable to all textures, irrespective of whether they are formed by a repeating element placed on the scene according to a regular grid.

We show that humans are able to assess texture scale irrespective of its regularity. In particular, we show that human assessment is consistent within and between subjects, and across methods of probing.

We measure human assessment of texture scale for the Brodatz dataset and provide estimates of the variability of each of the estimates. We make this dataset publicly available so as to enable researchers to make use of it as ground truth in the validation of other candidate methods and other applications.

We develop an algorithm for estimating texture scale which is validated

against human perception, and show that experts find the output of the algorithm useful on images from a wide range of disciplines.

We make our research available by means of an online calculator to which researchers and the wider public can upload images of textures and obtain estimates of their texture scale.

1.3 Thesis Outline

In Chapter 2, we present the relevant literature on texture. The purpose of this section is *(i)* to demonstrate the fit of our research in the realm of the existing literature, and *(ii)* to provide the reader with a concise introduction, from a basic level, to all relevant aspects of this work. Specifically, we present an overview of the relevant literature on "texture scale". Scale is highly important in image analysis, and as a result, the property of "scale" has been used to refer to many different things in the image analysis literature. We present some of the most commons definitions of texture scale, discuss how one relates to another, and review in detail certain methods in the literature for quantifying texture "scale", when used in a similar context to our characterisation for "scale".

In Chapter 3, we present two psychophysics experiments which we designed in order to *(i)* verify whether people are able to assess texture scale consistently, and *(ii)* measure human assessment. We then compare the two experiments and show that the results were highly consistent. In doing so, we demonstrate that humans perceive texture scale consistently for textures of all regularities. Furthermore, we show that the perception of texture scale is also consistent across different methods of probing. This chapter also provides the perceptual measurements of texture scale for all the textures in the Brodatz dataset and their corresponding confidence bounds. This can be used by other researchers to validate candidate models for the quantification of texture scale against human perception.

In Chapter 4, we present an algorithm for the quantification of texture scale. We also present a simplified version of the algorithm with greater computational tractability. We compare the performance of our algorithm to that of others in the

literature, and a selection of physical scales (related to texel size). We find that our method is able to predict human assessment of texture scale with a root mean square error (RMSE) of 1.2 just noticeable differences (jnds) on the Brodatz dataset (Brodatz, 1966b) (which contains textures of varying regularities). For a subset of regular and near-regular textures within the Brodatz dataset, our algorithm is able to predict human assessment of texture scale to within 0.7 jnds. A prediction with an RMSE of 1.0 jnds indicates that the sensitivity of our algorithm to texture scale is comparable to that of the visual system.

Given that this research was funded through an interdisciplinary research programme, we believe that for it to be complete, it needs to be made available to the research community. In Chapter 5, we present an online tool that we have created to give researchers direct access to our work and we also present the feedback that we have received from a test group which confirms that our predictions are reasonable and deemed to be useful.

In Chapter 6, we conclude the thesis with a summary of our contribution, a critical evaluation of it, and a proposal of what we believe the most natural extension of this work would entail.

1.4 Publications and Dissemination

The findings of this work have been published or presented in the following ways:

1. Live technology demonstration at UCL CoMPLEX 2014 Cumberland Lodge Retreat - May 2014

2. Oral presentation at the BMVA Student Symposium - March 2015

3. Poster presentation at UCL CoMPLEX 2015 Cumberland Lodge Retreat - May 2015

4. Poster presentation at ViiHM Meeting, Bath - July 2015

5. Oral presentation at ICSIA, Liege (Belgium) - July 2015

6. Extended conference abstract published in conference proceedings - Bouremoum S., Protonotarios E. D., & Griffin L. D. (2015). Quantifying texture scale. In *Acta Stereologica. Proceedings of the 14th International Congress for Strereology and Image Analysis.* Liege (Belgium).

7. Oral presentation at UCL CoMPLEX 2015 Cumberland Lodge Retreat - May 2016

8. Full length publication at IVC. First draft submitted 01 September 2016, corrections submitted 09 May 2017, currently under review.

Chapter 2

Background Material

Having provided an introduction to our work in Chapter 1, we now provide an overview of the relevant literature for the research contained in this work. We start by presenting an overview of the literature on texture scale and regularity in Section 2.1 where we will discover that there are no definitions of texture scale which are applicable to textures of all regularities. This will motivate our proposal of a new perceptual characterisation. We will then follow with a brief introduction to measurement theory and psychophysics, in Sections 2.2 and 3.3.2 respectively, which will set the framework for our approach to measuring human assessment of texture scale in Chapter 3. Next we will present an overview of texture databases (in Section 2.4 and a brief taxonomy of texture methods (in Section 2.5). The purpose of introducing these methods is to justify our choice of database, representation and approach. We continue to review in detail the methods of Ardizzone et al. (2013) and Hong et al. (2008) in Section 2.6 against which we compare the performance of our proposed algorithm. Lastly, we provide a brief summary of the key takeaways from our literature review in Section 2.7.

2.1 Texture, Scale, Regularity and Periodicity

We encounter texture everywhere, and while it is easy to perceive texture, there is no consensus as to how it should be defined, so authors often choose to define it according to the application at hand. Rosenfeld et al. (1982) define texture as a "similarity grouping in an image", Castellano et al. (2004) define texture as "the

appearance, structure and arrangement of the parts of an object within the image", while Sklansky (1978) states that "a region in an image has constant texture if a set of local statistics or other local properties of the picture function are constant, slowly varying or approximately periodic". Tamura et al. (1978) "regard texture as what constitutes a macroscopic region. Its structure is simply attributed to the repetitive patterns in which elements or primitives are arranged according to a placement rule". Lizarraga-Morales et al. (2011) treat texture as "a conjunction of two components i) a texture element (texel) and ii) a set of rules for texel placement into the field of view", while Hawkins (1970) argues that "the notion of texture appears to be dependent upon three ingredients: (i) some local 'order' is repeated over a region which is large in comparison to the order's size, (ii) the order consists in the non-random arrangements of elementary parts, and (iii) the parts are roughly uniform entities having the same dimensions everywhere within the textured region". A comprehensive collection of definitions for texture can be found in Coggins (1983). However it is defined, the ubiquity of texture in the visual perception and image analysis research is clear evidence of its importance. A fundamental reason behind the importance of texture in the image analysis field, is that its multiple attributes and properties have been shown to be useful in applications such as image segmentation, image retrieval, and object identification amongst many others. Some of these attributes, such as coarseness, regularity and isotropy, are perceptual (Rao and Lohse, 1993, Tamura et al., 1978) and have been studied psychophysically. Some have had algorithms proposed for their quantification. We are interested in the attribute of texture "scale"; both its perceptual assessment and its quantification.

The scale of texture is intricately related to its regularity and periodicity. In order to fully appreciate the limitations of the most common characterisations and quantification approaches for texture scale, we must begin by looking at the regularity and periodicity of texture. The most widely adopted paradigm for texture is that it is composed of two elements, *(i)* a repeating unit, and *(ii)* a placement rule. The study of texture regularity can be broken down into two main aspects. The first is the study of the regularity (or consistency) of the repeating texture elements (i.e.

an analysis of how similar texels are to one another). The second is a study of the regularity of the placement rule (i.e. measuring how array-like the texels are placed on the scene). This is often called periodicity. Before we introduce the concept of scale, it is beneficial to discuss the concept of texture regularity, to which texture scale is closely related, and which has also been comprehensively studied and characterised.

Nadler and Smith (1993) categorise texture as either regular or irregular, while others argue that there is a continuum of degree of regularity (Efros and Leung, 1999, Liu et al., 2004b). For the purposes of this work, we consider four levels of regularity: regular, near-regular, homogeneous and irregular. Regular textures are formed through the arrangement of a repeating unit called a primitive or fundamental tile, which appears in its original form or is isometrically transformed (Conners and Harlow, 1980, Schattschneider, 1978). Near-regular textures are often thought of as geometric, photometric or statistical departures from regular textures where the repeating elements can be easily identified (Lin et al., 2006a, Liu et al., 2004b). Regular and near-regular textures together can be referred to as periodic textures. Homogeneous textures are those where there is a similar repeating element but such element is not periodically placed on the scene (Ardizzone et al., 2013, Stam, 1997). In texture synthesis these are sometimes referred to as aperiodic textures. Figure 2.1 shows some example textures categorised according to their regularity.

The concept of scale is found throughout the image analysis literature. It has been used to refer to the resolvable detail in the image (Koenderink, 1984), the spatial resolution covered by each pixel (Marceau et al, 1990), the size of salient regions (Kadir and Brady, 2001), or the scale of the underlying filters in image representations (Griffin et al., 2009, Ojala et al., 2002b). Lindeberg (1994) noted that although the findings are very important they do not provide a method for choosing the most appropriate scale for further analysis.

Research on scale has focused on identifying scales of interest (scale selection, scale identification and saliency), integrating information across multiple scales (multi-scale methods), and on avoiding having to deal with scale selection (scale

Figure 2.1: Regular, near-regular, homogeneous and irregular textures (left to right).

invariance). Scale identification methods can be further split into methods that aim to identify global scales of a texture (Ardizzone et al., 2013, Bouremoum et al., 2015), and those that aim to identify local scale. Local scale methods assign a value to each pixel according to the scale of a neighbourhood centred at the pixel. Local scale maps are useful for segmentation based on textural features (Brox and Weickert, 2004, Hong et al., 2008).

A lot of work has been done trying to identify repeating elements and the corresponding arrangement lattice. Leung and Malik (1996), Schaffalitzky and Zisserman (1999), and Turina et al. (2001) propose methods for detecting repeating elements in a scene. These algorithms first aim to identify distinctive elements, and then search for repetitions in the image, taking into account planar transformations, such as perspective. They highlight that these could be texels on a texture, or more generally an item that appears multiple times in an image, such as faces in a photograph showing a group of people. Park et al. (2009) propose a three-step method for identifying the texton in regular and near-regular textures. They first extract interest points, which they use in the second stage to propose a parallelogram-shaped lattice unit. The second step in their method differs from that of Hays et al. (2006) in that the proposed lattice unit is computed by consensus of all pairs of candidate principal directions, instead of using a single location. Both methods then use the

Mean Shift Belief Propagation algorithm proposed by Park (2010) in order to refine the texel locations predicted using the two-principal directions calculated during the first step. Having found a partial lattice, both use a thin-plate spline wrapping method to expand the search for the repeating texel all the way to the edge of the image.

There are numerous applications associated with these methods. Researchers have used the spatial-structure of repeating elements in urban images to identify where in the city the image was taken (Schindler et al., 2008, Torii et al., 2013). Others segment textures based on the spatial arrangement of different texels in an image (Todorovic and Ahuja, 2009). Setti et al. (2017) propose an algorithm for counting repeating instances in an image. They first ask the user to manually locate a few instances of the repeating element. Their algorithm then computes three bag-of-word representations for each of the repeating elements selected by the user, and compares the representations across the selected repeats to identify a set of key features which are then used to look for other affine transformed repetitions of the texel in the image. Forsyth (2002) and Lobay and Forsyth (2004) make use of the spatial positioning of repeating elements in objects covered by a regular texture (such as checked shirts or polka dot dresses) to infer the 3D structure on the object being covered.

Specifically in relation to texture, scale is considered to be one of the most important features (Brox and Weickert, 2004). It is therefore not surprising that there is an abundance of work on scale in the image analysis and vision literature. Malik et al. (2001) define texture scale as the average distance between neighbouring texels. The theme of "repeating elements" is prevalent in scale definitions when refering to texture. These are all aligned with the texture paradigm of *a repeating element put on a surface according to a placement rule of some sort*. The primary limitations of defining texture this way are that it not only implicitly assumes a paradigm for texture, but such definitions are not suitable for textures with no discernible repeating elements. The way that Malik et al. (2001) define texture scale also requires that the repeating element is placed periodically on the scene and is

therefore also ill-defined for homogeneous textures.

Lin et al. (1997) use Gaussian smoothed autocorrelation functions together with maximum likelihood in order to identify centres of primitive patches. They then infer two vectors defining a parallelogram shape which they take to be the true primitive.

A co-occurrence matrix (CM) C, for an image I is a square matrix giving the relative frequency histogram of co-occurring pixel-values. For an 8-bit grey-scale and $m \times n$ image I, and for a δx-pixel horizontal offset and a δy-pixel vertical offset, the 256×256 ($= 2^8 \times 2^8$)CM is given by:

$$C_{\Delta x,\Delta y}(i,j) = \sum_{x=1}^{n} \sum_{y=1}^{m} \begin{cases} 1, & \text{if } I(x,y) = i \text{ and } I(x+\Delta x, y+\Delta y) = j \\ 0, & \text{otherwise} \end{cases}$$

The CM matrix can be computed for a range of $(\delta x, \delta y)$ pairs. Statistics, such as the (Pearson) χ^2 and κ (Cohen, 1968, Cohen and others, 1960), measuring agreement between rows and columns of C, can be calculated:

$$\chi^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{(C_{ij} - C_{i\cdot} C_{\cdot j})^2}{C_{i\cdot} C_{\cdot i}}$$

$$\kappa = \frac{\sum_{i=1}^{n} (C_{ii} - C_{i\cdot} C_{\cdot i})}{1 - \sum_{i=1}^{n} C_{i\cdot} C_{\cdot i}}$$

Parkkinen et al. (1990) compute a *measure of agreement*, κ (Cohen and others, 1960), from co-occurrence matrices for a range of displacements. Periodicity is determined by inspecting the agreement *vs.* displacement curves. A similar approach is adopted by Starovoitov et al. (1998), with the main difference being that they binarise the texture before looking for periodicity. This reduces the co-occurrence matrix for a given displacement from 256×256, for a typical 8-bit grayscale image, to a more manageable 2×2 matrix. Oh et al. (1999) derive an agreement measure that, for a given displacement, is equivalent to the *inertia* (Conners and Harlow, 1980, Peckinpaugh, 1991) of the co-occurrence matrix. In being able to compute the agreement without having to explicitly determine the co-occurrence matrices,

they show that their methods result in similar performance at a much reduced computational expense.

Structural methods based on autocorrelation functions include Lin et al. (1997) and Jan and Hsueh (1998). Both attempt to find the repeating texture primitive using autocorrelation functions and pixel gray-scale values. Lin et al. (1997) only consider regular textures, whereas Jan and Hsueh (1998) approach works for both regular and near-regular textures. Leu (2001) proposed a method based on the gradient field of textures, which encode the variation in intensity values at each point. Instead of simply searching for regular patterns of pixel intensities, Leu reports two separate measures to determine regularity of textures: the similarity between texels and the regularity of the texel locations on the texture. Liu et al. (2004a) identify the repeat element and the symmetry rules governing its placement. To do so they use autocorrelation functions, together with the definitions of the allowable transformations in Frieze and wallpaper-group patterns.

Park et al. (2009) propose a MRF method based on belief-propagation to identify deformed texels of regular patterns in natural images. Lin and Liu (2007) use a similar method for the identification and tracking of repeat units of deformed regular textures in video. Grigorescu and Petkov (2003) find the size of repeating units in regular and near-regular textures. They do so by identifying the window size giving the smallest entropy, which corresponds to the window size with the smallest number of visual patterns.

Near-regular textures are often treated as departures from regular textures. Ardizzone et al. (2013) propose a method to detect the global scale in regular and near-regular textures. Like us, they define the scale of the texture as the size of a repeating unit. They begin by finding keypoints on the image. They try Scale Invariant Feature Transforms (Lowe, 2004) (SIFT), Speeded Up Robust Features (Bay et al., 2006) (SURF) and Harris corner detectors (Harris and Stephens, 1988) and choose to use the keypoint detection method used in SIFT as this yields the best performance. For windows of a range of sizes, they compute the modal number of keypoints found in a random window of that size. They argue that there ought to

be a linear relationship between the modal number of keypoints and window size for windows of the repeat size and larger. Their scale estimate is therefore the point at which the curve becomes linear. To validate their model they manually outline texels in regular and near-regular textures.

We have found two methods in the literature that are applicable to a wider range of textures. Ardizzone et al. (2013) define texture scale as the smallest region for which the distribution of image keypoints becomes uniform (Ardizzone et al., 2013). It is claimed that this method is applicable to homogeneous textures as well as to regular and near-regular ones. Hong et al. (2008) define texture scale according to the size of the smallest domain size for which the distribution of grayscale values within this domain is similar to that of neighbouring domains. The authors claim that this method works for textures of all regularities. We present these two methods in detail in Section 2.6.

2.2 Measurement Theory

In this work we will attempt to characterise the scale of texture as a measurement process. There are two aspects to any measurement process that need to be addressed; the *representational* and the *pragmatic* aspects (Hand, 2004, 1996, Towsend and Ashby, 1984).

The representational aspect refers to the process of attaching numbers to objects in such a way that the numbers represent the empirical relationships between objects with reference to the attribute being quantified, in this case texture scale. Examples of such relationships include $A - B$,$a > B$, or $A + B = C$ where $+$ symbolises the concatenation of objects A and B. Depending on the structure of the relationship between the different objects, different types of scales, each with different properties, may be used. There are four main types of scales (Stevens, 1946, 1951):

- *Nominal scale:* A scale in which the numbers associated with objects act as labels. Only relations of equality may be made with these scales.
- *Ordinal scale*: A scale in which the numbers represent the rank of the ob-

jects. This type of scale is invariant to any order-preserving transformations. Relations of equality and rank-order may be made with these scales.

- *Interval scale:* A scale in which equal distance between the numbers associated with objects correspond to equal differences between objects. Interval scales have no absolute zero, and are invariant under linear transformations. Relations of equality, rank-order and equality of intervals may be made with these scales.

- *Ratio scale:* A ratio scale has the same property as an interval scale but comprises a well defined absolute zero. Relations of equality, rank-order, equality of intervals and equality of ratios may be made with these scales. Ratio scales are invariant under affine transformations.

The pragmatic aspect refers to the notion that the measurement process itself will define the attribute being measured. One may measure the same attribute in different ways. When the definition of the attribute being measured is open to interpretation, such as our perceptual characterisation of texture scale, it is important to check that the resulting values obtained represent the attribute accurately.

Perceptual estimates obtained through psychophysical experiments quantifying the magintude of an attribute in stimuli placed on scale can constitute a full measurement process as defined by Hand (1996), and thus comprise two aspects, the representational and the pragmatic. According to the representational aspect, the collection of data from humans in the psychophysical experiments is a way of identifying the relationship between the scale of textures.

2.3 Psychophysics

Fechner introduced the idea that perceptual intensity could be measured using experimental techniques as a means to relate the physical to the psychological in 1860. Since then a lot of work has been done in assessing human perception of the attributes of objects. These attributes are represented numerically and studied analytically, just like any other physical quantities such as weight or distance (Luce and Krumhansl, 1988).

A large number of psychophysical methods are available in the literature, and they can be categorised in a number of ways. One way to categorise psychophysical methods is into "threshold" approaches *vs.* "scaling" approaches. Threshold approaches aim to measure the limits of sensitivity. Scaling methods are one aiming to quantify stimuli intensities, and place them on a numerical scale.

Threshold approaches include the "method of adjustment", "method of limits" and method of "constant stimuli". The key idea behind these methods, is that for an attribute of a stimulus to generate a sensation in an observer, the stimulus intensity needs to be above a minimum threshold. Threshold methods aim to identify the minimum intensity level of an attribute that generates a sensation. In a method of adjustment experiment, the subject is able to adjust the stimulus stimulus attribute intensity, and they are asked to either adjust the intensity until it becomes almost imperceptible, or until the stimulus they are adjusting becomes most similar to a reference stimulus.

Threshold methods suffer from response bias, e.g. subjects may pretend they can perceive the attribute being tested in the stimulus when they do not (Ehrenstein and Ehrenstein, 1999).

Psychophysical scaling methods were defined by Torgerson (1958) "procedures for constructing scales for the measurement of perceptual attributes". In psychophysical scaling, we assume that every stimulus displays a given attribute to an intensity level ϕ, and the set of all possible stimuli intensities live on a continuous scale. In psychophysical experiments, one measures the magnitude of the sensation ψ caused by the stimulus on a set of observers. In order to construct a perceptual, one must first map the sensation magnitude, ψ to the stimulus intensity ϕ, and then place the inferred intensities on a continuous scale.

Psychophysical scaling methods can also be categorised into "direct" *vs.* "indirect" approaches. The most common direct approaches include magnitude estimation and absolute magnitude estimation. In a magnitude estimation experiment, subjects are asked to assign a numerical value to describe the sensory magnitude of the attribute being measured for various stimuli (Stevens, 1953, 1955).

There are advantages of direct scaling methods, for example, the fact that the value of the attribute being measured can be obtained either directly, or through a power law transform, based on the observation of the subject (Gescheider, 1985, Stevens, 1975, Thomas, 1981). However there are also some considerable disadvantages. Subjects' observations in direct scaling method experiments have been shown to be sensitive to calibration stimuli (by means of reference patterns), to the order in which the stimuli are presented, and even the wording of the exercise(Gescheider, 1985). The primary consequence of this, is that it is not possible to determine what the spacing between the intensity levels of the attributes should be on a scale (Gescheider, 1988).

In order to overcome these, perceptual sensation must be measured through indirect methods (Gescheider, 1988, McKenna, 1985). Forced-choice methods are the most common indirect scaling methods. In forced choice experiments, subjects are presented with at least two stimuli, or stimuli groups, and are asked to choose one of them. They are not able to skip the question, hence the word "forced". Experiments in which subjects have to make a choice between two options, are called 2-alternative forced-choice experiments, 2AFC in short. Two of the most common 2AFC tasks in visual psychophysics are pairwise comparison, and pair-of-pairs comparisons. In pairwise comparison experiments, subjects are asked to rank in turn pairs of stimuli with respect to the size of the estimated magnitude of the attribute in question. In pair-of-pairs experiments, subjects are asked to judge whether two stimuli in a pair are more different than two others in another pair with respect to the attribute in question. Indirect scaling methods overcome the weaknesses of the direct approaches, and more importantly, do not require consistent transitivity in order to fit a scale.

An advantage of pair-of-pairs is that it requires fewer trials in order to fit an interval scale compared to pairwise comparison, as it providers a relative sense of distance between observations. The pairwise method relies on internal noise and gives estimates on a discrimination scale. The scale will be in in units of discrimination noise. A scale constructed using data from a pair-of-pairs experiment, esti-

mates appearance-based scales relying on appearance differences. It is worth noting that while indirect scaling methods do not suffer from the same biases as the direct-scaling methods, they *(i)* generally require a larger number of observations to reach a given level of confidence on the estimates for any one stimulus (David, 1963), and *(ii)* the derived scales are no longer of the ratio type, but rather of the interval type.

The two most common linear scaling methods for psychophysical paired data are the Thurstone (Thurstone, 1927), and Bradley-Terry (Bradley and Terry, 1952) models. The Thurstone models assume that each judgement made by an observer is a noisy realisation (normally distributed) of the true value, while the Bradley-Terry model assumes that the probability of judging that stimulus A is greater than stimulus B is simply equal to the noisy realisation of the magnitude of the attribute exhibited by stimulus A over the sum of the magnitude of the attribute exhibited by stimuli A and B. The effective difference between the two models is that the Thurston model assumes a *probit* functional form, while the Bradley-Terry model assumes a *logit* functional form.

One advantage of the Thurstone model is that the difference (or preference) between two stimuli is also normally distributed. There are five types of Thurstone model. The difference between them surrounds the assumptions on the distributions of each observations. The simplest, and our preferred, case, is the Thurstone Case V model. This model assumes that the variance of the assessments are homoscedastic, i.e. have equal variance, σ. Under this assumption, the difference D_{AB} between stimulus A and B is given by $\Phi^{-1}(f_{AB})$, where Φ^{-1} is the inverse cumulative Gaussian and f_{AB} is the proportion of times for which stimulus A was preferred to stimulus B (Thurstone, 1927, Tsukida and Gupta, 2011). If we assume that the probability of any given respondent to choose stimulus A over stimulus B is constant, the number of time that stimulus A is preferred over B follows a binomial distribution. Because each comparison between two stimuli is a random distribution, it is possible that two given stimuli result in different preferences. In order to account for this difference, we carry out a maximum-likelihood optimisation in order to fit the observed preferences on an interval scale (Mosteller, 2006). The

log-likelihood is given by:

$$L(M|f) = \sum_{A,B} f_{AB} \log(\Phi(\mu_A - \mu_B)) \tag{2.1}$$

It is possible that errors occur during an experiment, i.e. a subject making the wrong choice either due to lack of focus or due to an accidental click. In order to account for this, one can introduce an additional parameter to the optimisation, λ, called the lapse parameter. The introduction of the lapse rate accounts for the sensitivity of maximum likelihood methods sensitivity to frequencies near 0 and 1. Consider stimuli A and B such that $A >> B$, and that a careless subject accidentally judges $B > A$ in one of the ten comparisons in the dataset, reducing the frequency of time in which A was preferred to B from 100% to 90%. When fitting a sigmoid shaped Gaussian CDF, the ML will penalise the other preference by adjusting the point slope to capture the incorrectly low frequency. Including the lapse parameter. The mathematical effect of introducing the lapse parameter, is that the preference function Φ^{-1} gets rescaled from $[0,1]$ to $[\lambda, 1-\lambda]$, and therefore does not unfairly penalise correct response with incorrect responses near 0 and 1 (Swanson and Birch, 1992, Wichmann and Hill, 2001).

When designing a psychophysical study, the experiment needs to carefully consider the set-up and conditions under which the experiment will be carried out. In a controlled experiment, one can ensure that there are no differences in the display brightness, contrast, level of focus, etc. and that the distance to the display, and viewing angle are kept constant. Controlled-conditions have been favoured in experiments in which the attribute measured is sensitive to these features, for examples when assessing brightness (**?**). Running experiments online, in uncontrolled experiments, offers more convenience, speed and variability of respondents. These experiments are particularly useful for studies in which the results will be used in applications that will not take place under controlled experiments. An increasing number of experiments are being carried out in uncontrolled conditions, for example, in colour naming exercises (Mylonas et al., 2010). While having controlled conditions can be important in studies in which subjects are asked to make an ab-

solute judgement, it is less so in forced-choice experiments in subjects are asked to make relative assessments between multiple stimuli, as any pair, pair-of-pairs will be viewed under the same conditions.

2.4 Textures, Databases and Applications

In this section, we will review some of the most commonly encountered texture datasets in the literature. We will also examine a few applications in which these datasets are used.

In order to evaluate a candidate method for estimating texture scale, it is necessary to critically compare the candidate method to other methods in the literature and assess its performance using standardised datasets. Depending on the task at hand, or the desired properties of the method, one might want to evaluate a candidate method on more than one type of image. Fortunately, there are many texture datasets available to researchers on which such a critical evaluation and performance assessment can be based.

Hossain and Serikawa (2013) provide a comprehensive survey of common texture databases which is extended by Bianconi and Fernández (2014). Together, these reviews survey in excess of 45 such databases, and organise them according to a four-category taxonomy: (i) texture databases in bio-medicine, (ii) natural texture datasets, (iii) texture of materials databases, and (iv) dynamic texture databases.

Most databases organise images according to categories and/or label images according to their content. Some datasets contain one image per category (label) (these tend to be larger images) while others contain multiple images per category.

Although we adopt a similar taxonomy when presenting these databases to that adopted by Hossain and Serikawa (2013) and Bianconi and Fernández (2014), we note that some of the databases could fall into multiple categories. We further note that some of the databases could be categorised in different ways, for example, relative to the type of application for which they are used.

2.4.1 Natural Images

Natural images are images which show objects as they might be observed in the real world. Examples include images from nature, such as foliage, mountains, trees, seas and oceans, as well as built up environments, for example, images of people, buildings and satellite view images. Satellite images can be split into further categories, for example, databases for remote sensing applications.

Natural texture image datasets show textured surfaces from the real world and are popular for the evaluation of texture analysis methods. By far. the most well-known natural texture image dataset by far is the Brodatz dataset (Brodatz, 1966a). Originally intended as a repertoire for artists and designers, the Brodatz dataset is composed of 112 gray-scale images. The images are highly varied and suffer few illumination and contrast issues. Because of this, the Brodatz dataset has become the defacto dataset for testing texture analysis algorithms. Figure 2.2 shows twelve of the Brodatz textures.

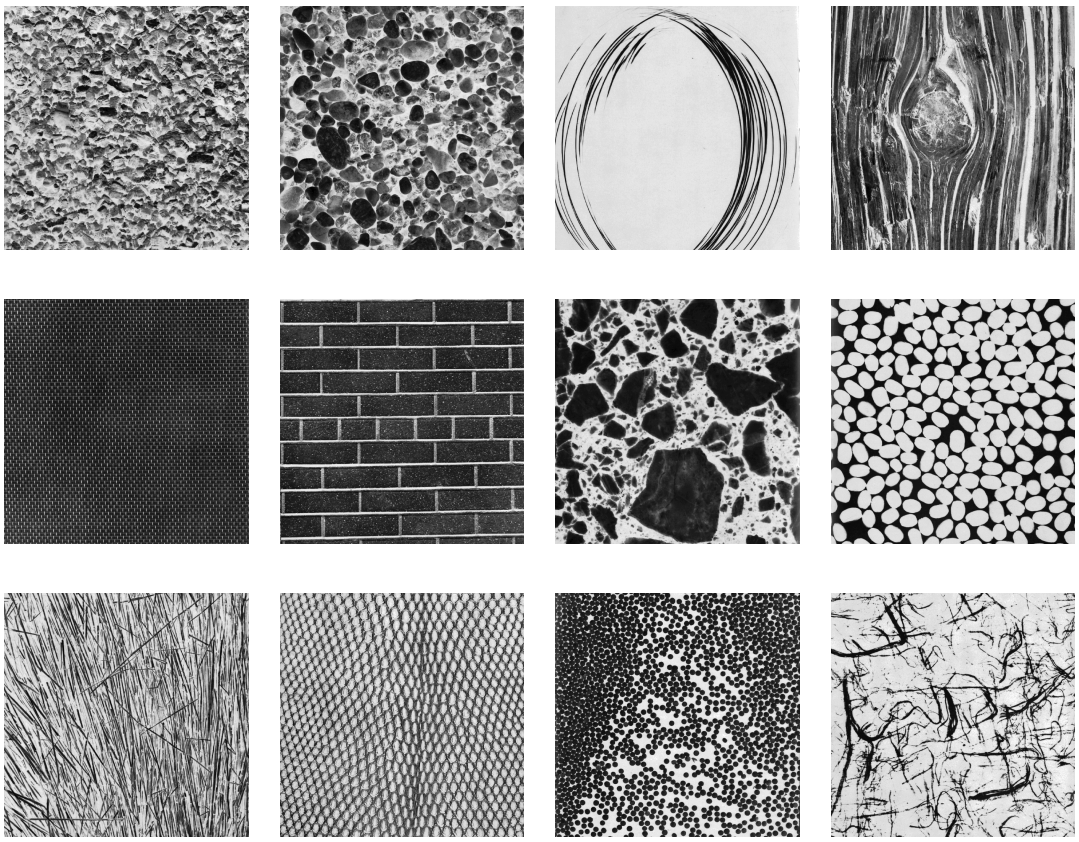


Figure 2.2: Twelve original Brodatz textures

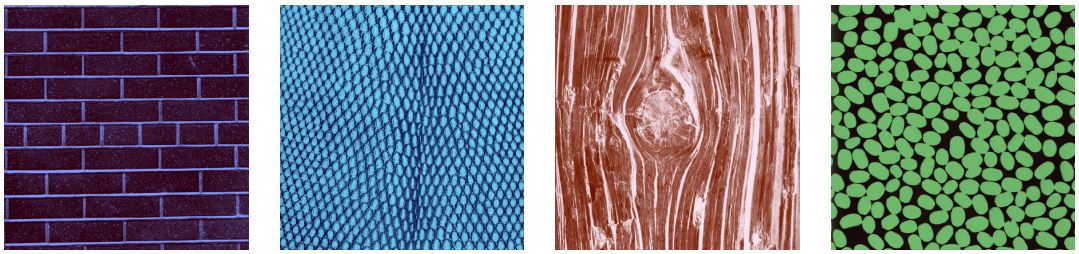


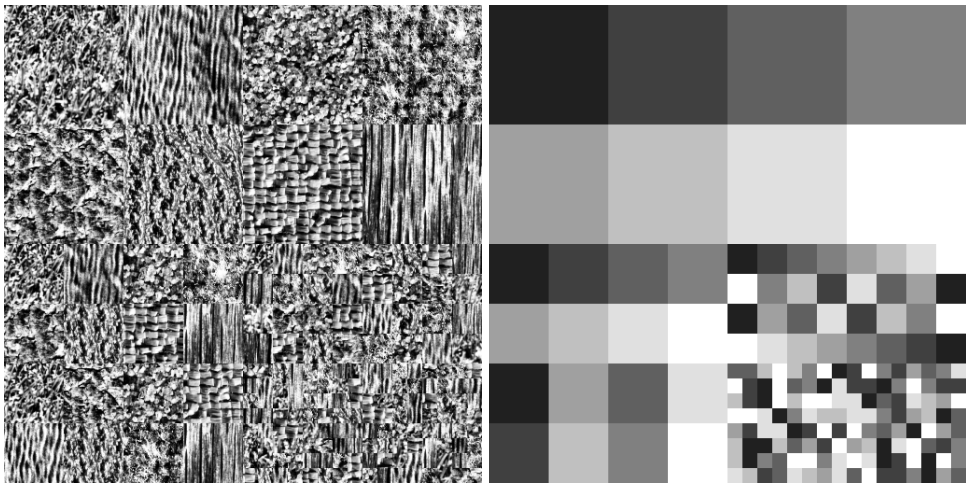
Figure 2.3: Four examples from the coloured Brodatz dataset (Abdelmounaime and Dong-Chen, 2013)

Ojala et al. (1996) use the Brodatz to compare the effectiveness of several texture measures for classification. The Brodatz has also been used to evaluate the effectiveness of image features in the proposal of new image features (Liao et al., 2009, Manjunath and Ma, 1996, Smith and Chang, 1994).

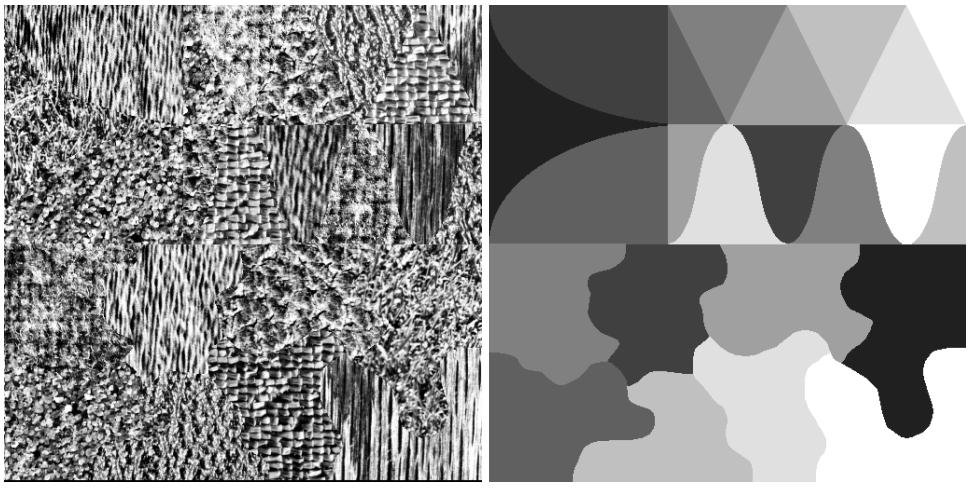
There have been attempts to synthesise the Brodatz (Efros and Leung, 1999, Paget and Longstaff, 1995, Wei and Levoy, 2000) and despite claims that the Brodatz is an easy dataset for retrieval, it is also widely used to test the performance of textural features in retrieval tasks (Kim et al., 2002, Ojala et al., 1996, Po and Do, 2003, Unser, 1995). Further, the Brodatz dataset has been coloured by Abdelmounaime and Dong-Chen (2013) so that it can be used to evaluate texture methods designed to work on RGB images (some of the coloured Brodatz are shown in Figure 2.3).

Because the Brodatz are taken in controlled light conditions, they suffer few defects (such as occlusions and background regions) and are very diverse. They are also very popular in testing texture methods which requiring test images with few non-textured regions. Although the database has been criticised by some as being an easy choice for retrieval tasks, others argue quite the opposite: that it is not an easy dataset to use due to its diversity (Lazebnik et al., 2005b).

Vistex (Picard et al., 2010) is another popular dataset containing 167 coloured images. The images are not controlled for perspective or lightning conditions so while they offer a larger intra-class variability than the Brodatz images, they also contain non-textured portions of images and images with more than one textured region per image. It is therefore considered more difficult to use in relation to



(a) Mosaic texture 2 from USC-SIPI database (left) with ground-truth segmentation (right)



(b) Mosaic texture 3 from USC-SIPI database (left) with ground-truth segmentation (right)

Figure 2.4: Two USC-SIPI texture tiles and corresponding segmentations.

retrieval tasks. Similarly to the Brodatz, Vistex is also widely used for retrieval (Do and Vetterli, 2002, Kim et al., 2002, Permuter et al., 2003).

USC-SIPI (Weber, 1997) comprises four texture mosaics and their ground-truth segmentation. This dataset is primarily used for segmentation (Hsiao and Sawchuk, 1989, Lin et al., 2006b, Wassenberg et al., 2009), edge-detection (Baştürk and Günay, 2009), and image encryption (Pareek et al., 2006, Wu et al., 2012).

A more modern and unusual database is the "Mayang's free textures" dataset. It comprises over 4350 images grouped into 9 classes, each of which is then divided further into sub-categories (Murni Adnin and Smith, 2001). Some of the images display constant texture (similar to the Brodatz), while others contain scenes

with more than one textured region. The high variability within the classes and the non-homogeneity of many of the images make it a challenging dataset to use for classification tasks (Kumar et al., 2016, Ramakrishnan and Selvan, 2008).

Other natural texture databases discussed in the review by Bianconi and Fernández (2014) include the Salzburg texture image dataset (STex) (Roland and Meerwald, 2009), which contains 476 images taken around the city of Salzburg, and USPTex which comprises 191 images of general scenes, such as textiles, walls, vegetation and roads (Backes et al., 2012).

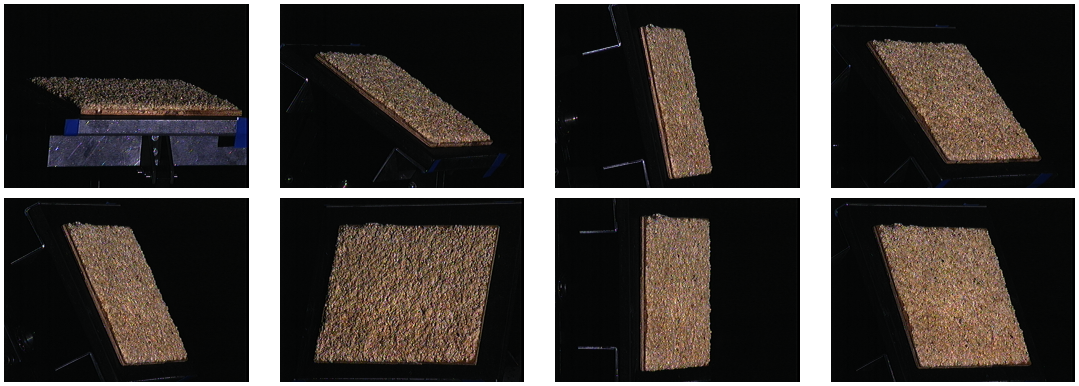
2.4.2 Material Textures

Material texture datasets can be similar to some of the natural texture datasets. Like natural image datasets, they often contain multiple images for each class, they are popular in texture classification tasks. Material texture datasets are the most prevalent type within the world of texture databases.

The CUReT (Columbia-Utrecht Reflectance and Texture) database comprises images of 61 materials (classes). Each material has been photographed from a range of orientations but with a single light source. There is a black background around most samples and the mechanism/stand on which the photographed materials are placed is visible in some of the images (Dana et al., 1999). The large number of samples per class makes this a good database for classification tasks and it is used by a large number of researchers (Guo et al., 2010, Hayman et al., 2004, Liu and Fieguth, 2012, Varma and Zisserman, 2003, Zhang et al., 2006). Figure 2.5 shows some examples from the CUReT database.

The KTH-TIPS (KTH Textures under varying Illumination, Pose and Scale) database (Fritz et al., 2004) is composed of 10 classes. Each class contains 81 images taken with varying lighting directions (3 Illumination), angles (3 Poses) and magnification (9 Scales) ($3x3x9 = 81$). The KTH-TIPS has an advantage over CUReT in that the images have very little to no non-texture borders. Because of their variety in the IPS dimensions, they are very popular for feature identification and classification tasks (Crosier and Griffin, 2010, Hayman et al., 2004, Nowak et al., 2006, Zhang et al., 2007). Figure 2.6 shows some examples from the KTH-

(a) Eight distinct CUReT materials



(b) Eight distinct views of a CUReT texture

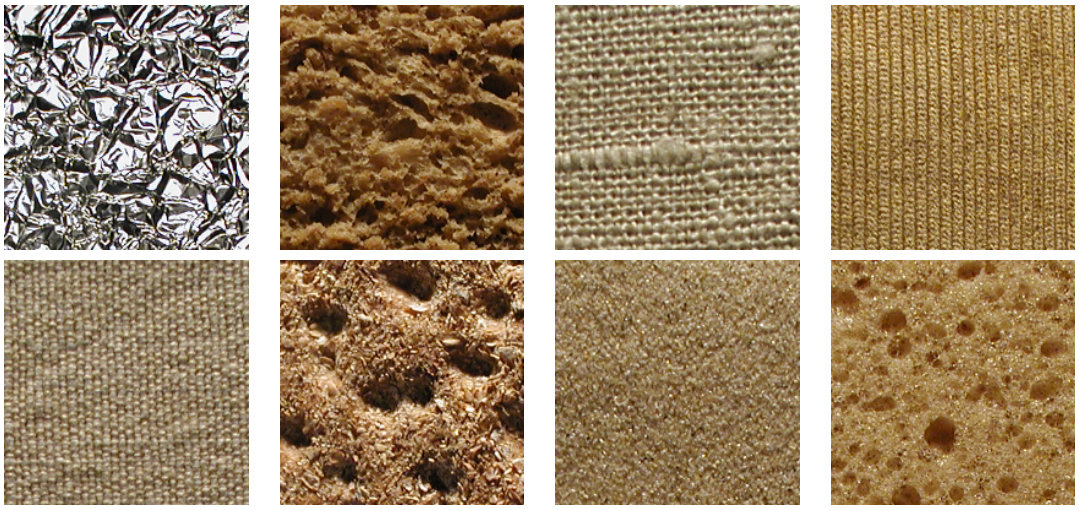
Figure 2.5: Eight CUReT textures (a) and eight views of CUReT sample 08-01

TIPS texture database.

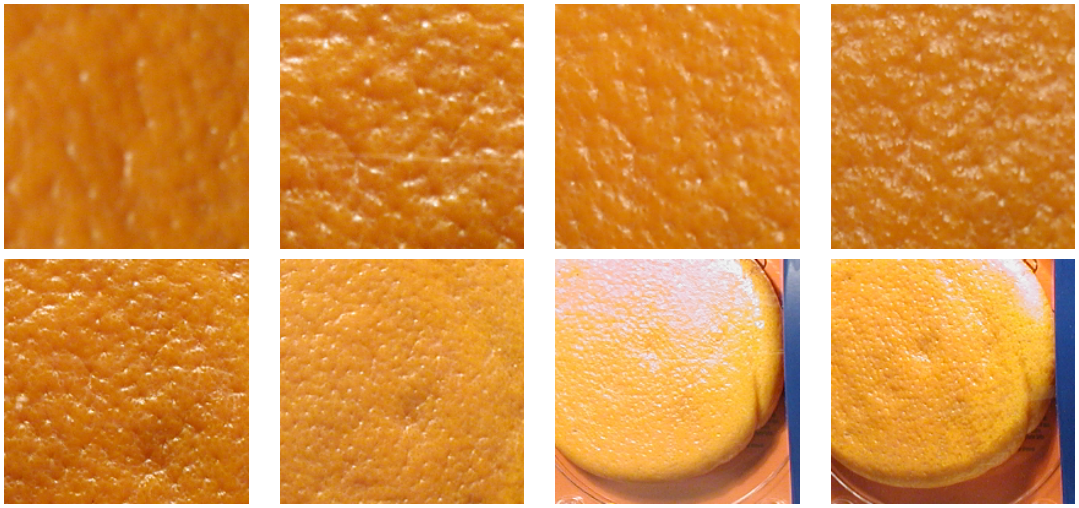
KTH-Tips2 extends KTH-TIPS to include 432 samples for 11 texture classes. Mallikarjuna and Targhi (2006) provide details on the Illumination, Pose and Scale variation used for each class, as well as details on post-processing, such as cropping boundaries and the methods used for dealing with poor quality images.

Photex contains images of the same surface taken at 40 different illumination directions (Various, 1999). It is used for classification tasks in Drbohlav and Chantler (2005), Targhi et al. (2008). There is also a 3D version of the Photex which includes real surface rotation as opposed to image rotation. Photex comprises 1680 images of 30 textured materials. This dataset is popular with shape from texture methods (Dong and Chantler, 2004, Dong et al., 2007, Jian and Dong, 2011). Figure 2.7 shows some examples from the Photex dataset.

The UMD (University of Maryland Dataset) dataset contains 25 classes of im-



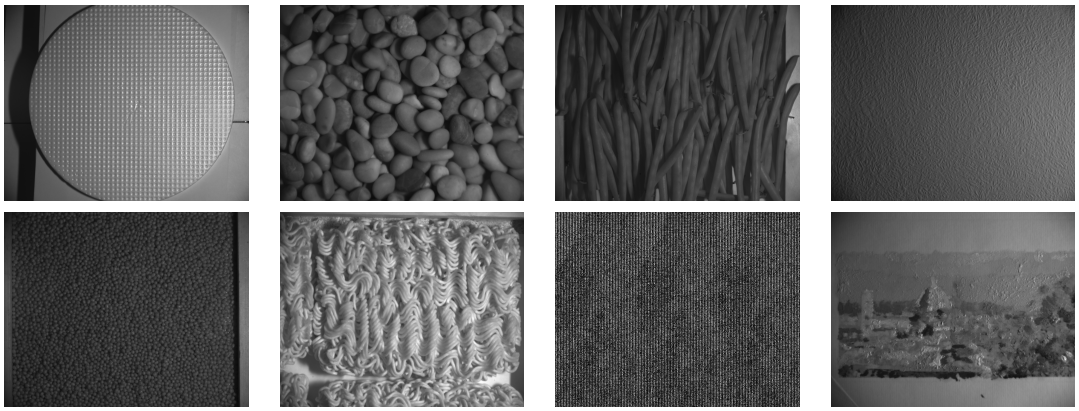
(a) Eight distinct KTHTips materials



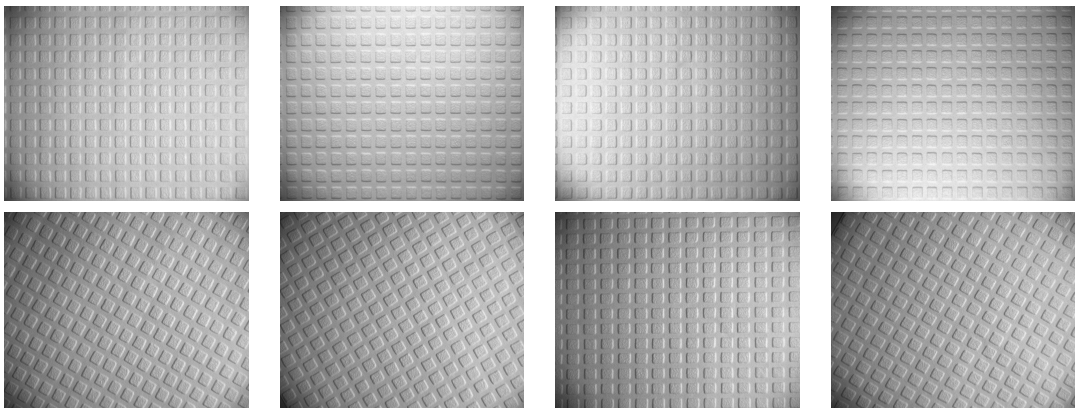
(b) Eight distinct views of orange peel from KTHTips database

Figure 2.6: Eight KTH-Tips textures (a) and eight views of orange peel from KTH-Tips

ages with 40 gray-scale images per class of varying viewpoints and zoom levels. This is quite a challenging dataset because the illumination conditions are not controlled and also because the images have been taken using a non-calibrated amateur camera (Various, 2006). Image categories include several types of bark, wood, brick, carpets, textiles, stone, as well as water. The fact that each image shows a unique texture and the classes of image are well curated, make this a very good dataset for both identifying texture features (de Siqueira et al., 2013, Xu et al., 2012, 2010) and evaluating their performance in classification tasks (Ji et al., 2013, Kong and Wang, 2012, Liu et al., 2011, Quan et al., 2014).



(a) Eight distinct PhoTex materials



(b) Eight distinct views of one PhoTex image

Figure 2.7: 8 PhoTex textures (a) and eight views of one PhoTex texture.

The Outex dataset (Ojala et al., 2002a) includes both natural images and material textures. This dataset contains images of 320 textured surfaces organised into 29 classes. The dataset is very large as each of the 320 textured surfaces was imaged under three different types of light source, nine rotations and six different resolutions. There are therefore 51,840 samples, in both RGB and gray-scale. Like many other material datasets, it is primarily used to evaluate the performance of texture features in classification tasks (Arvis et al., 2011, Mäenpää and Pietikäinen, 2004, Ojala et al., 2002a,b, Ojansivu and Heikkilä, 2008).

The UIUC database (Lazebnik et al., 2005b) contains 40 gray-scale images for each one of 25 classes, with variation in magnification and viewpoint. This dataset also suffers variation due to uncontrolled illumination conditions. It is considered a challenging database for classification due to the large variations in viewpoint

and magnification in the 40 images per class (Varma and Garg, 2007). The UIUC dataset is most often used in evaluating the performance of textural features for classification (Lazebnik et al., 2005a, Mellor et al., 2008, Sifre and Mallat, 2013, Zhao et al., 2012).

The MEASTEX (MEASurement of TEXture) database contains a combination of images of homogeneous-natural and artificial materials (Smith and Burns, 1997). They are divided into three broad categories, namely grass textures, materials textures and surface textures. The grass category contains different types of grass in a range of presentations. The materials category contains images of various classes of mulch, sandstone, gravel, pebbles and sand. The surfaces category is divided into various types of asphalt, concrete, corrugated iron and office partitions.

There are many other material databases available in the literature, such as the The PSU-NRT (PSU Near Regular Texture) dataset (Liu et al., 2004b) and the ALOT (Amsterdam Library Of Textures) (Geusebroek and Burghouts, 2009) which are diligently reviewed in Hossain and Serikawa (2013) and Bianconi and Fernández (2014).

2.4.3 Texture in the Biosciences

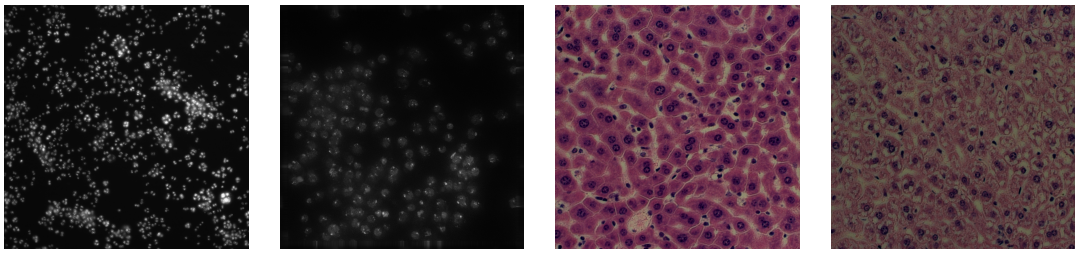
With the improvement of medical imaging over the last few decades, researchers are able to see more inside living creatures. Images of cellular tissues have been joined by others from X-Ray, MRI, and CT scans in the medical image analysis fields. Many texture analysis methods have been used in bio-medical libraries for the identification of (segmenting) organs and other body parts: organs in the abdomen (Koss et al., 1999), bone structures (Lorigo et al., 1998), knee (Sebastian et al., 2003), carpal and hand (Sotoca et al., 2003) amongst many others.

Texture analysis methods have also been used for cell segmentation (Kong et al., 2011, Sharma et al., 2008) and identifying tumours: malignant liver tumours, (Xian, 2010), breast tumours (Chen et al., 2002) and colon carcinoma (Esgiar et al., 2002).

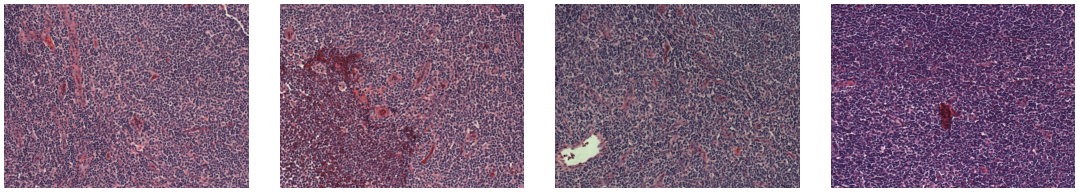
Any image bank with pictures of cell cultures can be treated as a texture database. Here we present a few large and publicly available datasets and show

some images from these datasets in order to illustrate texture methods used on such images.

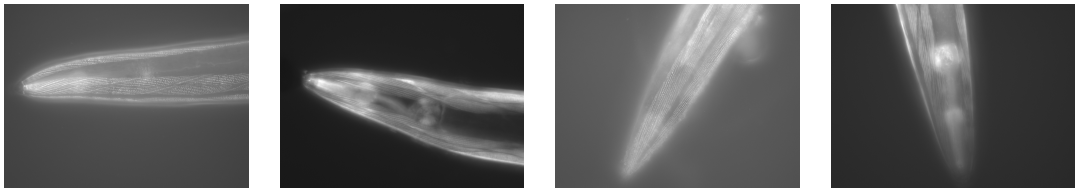
A large and popular dataset in the biosciences is the IICBU biological image repository. This dataset was proposed to be used as a benchmark dataset for comparing the performance of image analysis algorithms for biological images. It contains 11 libraries of images. Figure 2.8 shows some example images from the IICBU library.



(a) Some samples from the IICBU dataset



(b) Lymphoma cell tissues from the IICBU dataset



(c) Head of C Elegans roundworm 1, 2, 4 and 8 days after hatching

Figure 2.8: Examples from the IICBU Biological Image Repository

MESSIDOR is a dataset containing images of eyes with retinopathy collected by three hospitals in France. Some examples are shown in Figure 2.9, and these images have been used to develop tools that automatically detect the disease (Acharya et al., 2012, Jayasakthi and Rajaselvi, 2016) as well as to detect the formation of new vasculature on the retina (Lee et al., 2013).

The USF database for screening mammographic mass (USF-DDSM) comprises mammography images which have been used in the development of tools

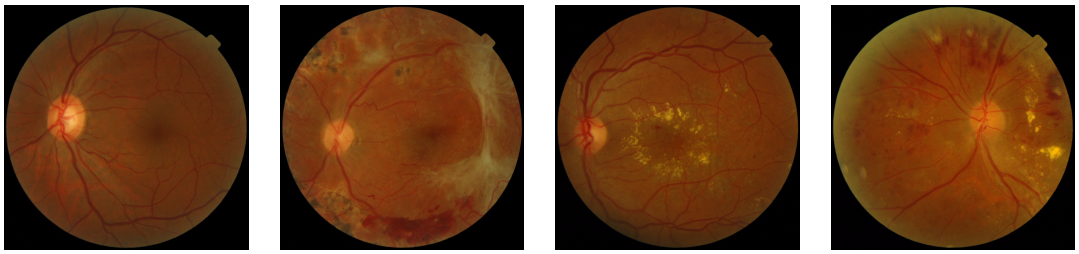


Figure 2.9: Four retina images from MESSIDOR

for estimating mammographic mass (Bruynooghe, 2006, Zhang et al., 2011, 2010) which helps clinicians identify regions of interest in the detection of suspicious formations and malignant masses within breasts (Wang et al., 2007, Zhang et al., 2012).

Bianconi and Fernández (2014) and Hossain and Serikawa (2013) present a few more datasets for skin conditions (Rutgers) (Cula and Dana, 2004, Cula et al., 2004), lung emphysemas (Sorensen et al., 2010), and epistromas (which show samples of bowel cancer) that can be accessed through an online interface. The common feature of all these datasets is that they have been used to develop and validate texture analysis methods.

2.4.4 Dynamic Textures

In the previous three sections we considered static images of static objects. Dynamic texture datasets refer to those in which the object is being photographed/recorded over time. In other words, dynamic textures are time-varying textures that are often distributed in video format. Each video frame can be taken as a sample of a class (Doretto et al., 2003a,b, Doretto and Soatto, 2003, Soatto et al., 2001).

We will not review dynamic texture datasets in great detail as we do not discuss them anywhere else in the work, but they have been included here as we feel that it is necessary to acknowledge their importance in certain texture analysis applications, especially in recognition and synthesis applications (Costantini et al., 2008, Efros and Freeman, 2001, Li et al., 2002, Yuan et al., 2004). The three most common datasets in dynamic texture analysis are DynTex, Dyn-Tex++ and the UCLA

Dynamic Texture Dataset, which we review briefly.

The UCLA dynamic texture dataset (Saisan et al., 2001) includes gray-scale images sequences for smoke, a fountain, a flowing river, ocean waves and a flowing curtain, as well as colour RGB sequences of a fountain, ocean waves, fire and small waves on stagnant water (Doretto et al., 2003a). This dataset has been useful in evaluating texture features on both texture synthesis as well as for use in classification and recognition tasks (Chan and Vasconcelos, 2007, Derpanis and Wildes, 2010, Ghanem and Ahuja, 2010, Ji et al., 2013).

DynTex (Peteri and Chetverikov, 2005, Peteri et al., 2010) is a popular comprehensive dataset of dynamic textures maintained by the university of La Rochelle. The creators intended for it to become the reference database for dynamic textures, in the same way that Brodatz is the reference dataset for natural textures. DynTex is certainly a large dataset: it contains over 650 sequences of not less than 250 frames per sequence, making the dataset over 52GB in size. It has been used for recognition tasks (Fazekas and Chetverikov, 2007, Zhao and Pietikainen, 2007) as well as for synthesis (Filip et al., 2006, Guo et al., 2013).

Ghanem and Ahuja (2010) have criticised the UCLA dataset for its small size and the Dyntex dataset for a lack of labelling in relation to the sequence, panning and zooming of the video (present within some of the sequences). Ghanem and Ahuja further criticised the fact that the static and dynamic background has not been cropped out, as well as the huge size of the data (as it was provided in RAW format). They decided to run some pre-processing on the DynTex dataset and produced a curated dataset containing 3600 sub-sequences (1000 in each of 36 classes), obtained by curating 345 of the DynTex sequences. None of the resulting sequences show more than one dynamic texture, they contain no background, no panning or zooming, nor do they include scenes with little or no motion. Ghanem and Ahuja named the dataset Dyntex++. It is publicly available and has been used in texture classification and recognition tasks (Päivärinta et al., 2011, Ren et al., 2013, Tiwari and Tyagi, 2017). For a further review, the reader is referred to the comprehensive work of Bianconi and Fernández (2014) and Hossain and Serikawa (2013).

2.5 A Taxonomy of Texture Representations

We present a brief summary of some of the most common texture methods arranged into four categories: structural, statistical, model-based and transform-based. Texture methods may fall into more than one category.

2.5.1 Statistical Texture Methods

Texture analysis methods often need to be applied on real-world, *natural*, images. In the simplest case this refers to images of a single, uniform, (*a priori*) textured surface. For a texture analysis method to be practical it often needs to be robust. Robustness is often taken to be a proxy for invariance against properties irrelevant to the task at hand. These are primarily manifested as grey-level variations within and between images of the same texture (class) due to perspective shifts, rotations, illumination and scaling variations. The human visual system is robust against these properties; when presented with a texture, changing the angle, light and size of the presentation will, in most cases, not affect our conviction that we are looking at the same texture (class).

Many statistical methods consist of describing an image region as an histogram of *features* found in said region. The standard process for this type of statistical method is as follows:

1. Choosing a set of filters to apply to an image;

2. Choosing a set of features from the filter responses;

3. Choosing how to represent the features as an histogram.

Methods also vary depending on whether or not the histograms across all textures are defined across the same alphabet. Methods based on feature distributions are called *bag of visual words* or *bag of textons* methods.

Ojala et al. (1996) and Ojala and Pietikäinen (1999) introduce the concept of *local binary patterns* (LBP), as a gray-level difference based simplification of Wang and He (1990) texture-spectrum local neighbourhood labelling method. An LBP is

given by:

$$LBP_{n_P,r} = \sum_{i=1}^{n_P} 2^{i-1}\delta(g(i) - g(0)), \tag{2.2}$$

where n_P is the number of equally spaced pixels at a radius r around the central pixel, $\delta(x) = 0$ for $x < 0$, $\delta(x) = 1$ for $x \geq 0$, and $g(i)$ are the gray-scale values of pixels with coordinates $\left(-r\sin(\frac{2\pi i}{n_P}), r\cos(\frac{2\pi i}{n_P})\right)$, given that the coordinates of the central pixel $g(0)$ are $(0,0)$. In their later paper (Ojala et al., 2002b), they also introduced a measure for the contrast of local texture regions, given by the variance of the local pixels:

$$VAR_{n_P,r} = \frac{1}{n_P}\sum_{i=1}^{n_P}\delta(g(i) - \bar{g})^2, \tag{2.3}$$

where $\bar{g}$ is the mean of the $g(i)_{i=1}^{n_P}$.

Ojala et al. demonstrate the rotation invariance and discrimination power of $LBP_{n_P,r}$, $VAR_{n_P,r}$ and the ratio $\frac{LBP_{n_P,r}}{VAR_{n_P,r}}$, via a series of classification tasks on the Outex database. The best classification performance, 97.9% is obtained using $\frac{LBP_{n_P,r}}{VAR_{n_P,r}}$. Ylioinas et al. (2013) propose a novel method for sampling LBPs, which they call *dense LBPs*. Their sampling method consists of taking internal pixel corners in an image as $g(0)$ (centre of their LBP samples), as well pixel centres, in standard LBP. For an $m \times n$ image, the standard, say, $LBP_{8,r}$ representation would contain $(m-r)(n-r)$ points. Ylioinas et al. (2013) equivalent LBP_8, r variant would contain $(m-r)(n-r) + (m-r-1)(n-r-1)$ points. This increase in density leads to slightly improved classification performance ($\approx 2-8\%$) compared to standard LBP. Liu et al. (2012) propose another extension to the standard LBP extension. In their method, they calculate four LBP variants at every pixel centre. Their first two LBP variants capture the intensity of the central and neighbouring pixels separately. The other two variants capture pixel value difference in angular and radial directions. They report a significant improvement ($\geq 8\%$) in classification performance compared to standard LBP.

Lowe (1999, 2004) introduces the scale invariant feature transform (SIFT) keypoint detection and representation. He uses a difference of Gaussian (DoG) detector to identify local extrema which he uses as *keypoints*. He then filters down the key-

point sample by first deleting low contrast points and then deleting keypoints for which the principal curvature is above a pre-set threshold, which he argues correspond to edges. Next, he defines an image representation based on the identified keypoints, using local gradients. He shows that this representation performs well in object identification tasks. Ke and Sukthankar (2004) propose a method, PCA-SIFT, to reduce the dimensionality of the SIFT representation and show that the performance in a retrieval task improves compared to the standard SIFT method. They also show that the computational cost in performing PCA on the SIFT representation is small compared to the saving in the matching cost in retrieval tasks.

Bay et al. (2006) introduce a set of features, *speeded up robust features* (SURF). Just like SIFT ,the features are based on local extrema keypoints. They introduce an extrema detector that uses the Hessian of the image convolved with a Gaussian kernel and show that, for an image of 800×640 pixels, their method finds a comparable number of keypoints to the DoG (1418 *vs* 1520) in under one third of the time. They then introduce an descriptor, based on Gaussian weighted wavelet responses, in horizontal and vertical directions. They show that their joint detector-descriptor is up to $4\times$ faster than SIFT and recall performance improves by up to 10%.

Leung and Malik (2001) propose a method for obtaining *textons* from textures. They apply a set of 48 DtG kernels to multiple images of the same surface under different illumination and viewpoint conditions. They argue that by doing this they are capturing both the geometric and photometric properties of texture, thus they call their vocabulary *3D textons*. They then concatenate the responses for all images and all filters. They obtain their feature vocabulary by clustering the filter response space (using k-means) and then joining tightly packed clusters for which there is likely to be little data.

Zhang et al. (2007) use the affine invariant version of the Harris-Laplace and Laplacian detectors (which capture corner-like and blob-like regions) together with the SIFT, RIFT and SPIN descriptors to construct their keypoint space. They learn a feature space by clustering their keypoints using a support-vector-machine (svm)

classifier. They report improved classification performance when combining the two detectors and three descriptors compared to using them individually on a range of texture datasets.

Griffin et al. (2009) introduce the Basic Image Feature (BIF) texture representation which we review in detail in Section 4.1.1.3.

Zhang et al. (2015) introduce a scale invariant texture representation based on frequency decomposition and gradient orientation. It consists of a 2D histogram for the joint distribution of orientation-decomposed image-intensities (responses to wedge filters) and a texture gradient. They compare classification performance on two texture databases to the performance of other texture representations. They find that their algorithm marginally outperforms ($\approx +0.3\%$) the state-of-the-art (BIFs) for one of the datasets, but performs considerably less well on another ($\approx -4\%$), compared to BIFs.

Other statistical methods include co-occurrence matrix methods (Clausi, 2002, Gotlieb and Kreyszig, 1990) and autocorrelation methods Haralick (1979), Ulaby et al. (1986).

2.5.2 Transform Based Methods

Campbell and Robson (1968) carry out a series of psychophysical experiments to probe human perception of gratings. They argue that the results of the experiments indicate that there are two separate perceptual mechanisms for frequency and orientation in the human visual system. Similar to Fourier decomposition in signal processing, the human visual system decomposes visual signals into multiple, frequency and orientation tuned channels (Georgeson, 1979).

The Fourier transform, $\hat{f}$, for a one dimensional signal f is given by:

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x \xi} dx \tag{2.4}$$

One can introduce spatial dependency into the Fourier representations by restricting it to a specific window. The *short-time* Fourier transform (STFT), $\hat{f}_w$, of a one dimensional signal f, in a window w is given by:

$$\hat{f}_w(\xi) = \int_{-\infty}^{\infty} f(x)w(x-\xi)e^{-2\pi ix\xi}dx \tag{2.5}$$

The STFT in the special case for which w is a Gaussian function is called a Gabor transform (Tuceryan and Jain, 1998). Features extracted from Gabor filters composed of trigonometric functions with Gaussian envelopes have been used for texture segmentation (Dunn and Higgins, 1995, Grigorescu et al., 2002, Jain and Farrokhnia, 1991, Weldon et al., 1996) and classification (Arivazhagan et al., 2006, Idrissa and Acheroy, 2002)

The wavelet representation W_f of a function f is its decomposition into an orthonormal series:

$$W_f(\xi) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(x)\psi(\frac{\bar{x-\xi}}{a})dx, \tag{2.6}$$

where $\psi_{kl}(x) = 2^{k/2}\psi(2^k x - l)$.

Features from wavelet transforms have also been successfully used in texture classification (Arivazhagan and Ganesan, 2003, Chang and Kuo, 1993, Portilla and Simoncelli, 2000) and segmentation (Unser, 1995) tasks. Wavelet representations are often preferred to Gabor ones because its components are orthogonal, so there is no redundancy in the extracted features (Materka et al., 1998, Teuner et al., 1995).

2.5.3 Model-Based Methods

Model-based texture methods model textures by a functional (not necessarily explicit) form. The models are parametrised by parameters that capture the essential properties of the texture. Model-based texture methods often consist of identifying the parameter values that lead to the most life-like synthetic copies of the original texture. Model performance is often assessed by comparing synthesised textures with the real ones. They are very common in texture synthesis methods. Common methods include Markov Random Field -based and Fractal methods.

A Markov Random Field is an undirected network structure that represents dependencies between random variables. Nodes in the network represent the random variables. Edges connect pairs of nodes for which a statistical dependency exists

(Li, 2012). When using MRFs for texture analysis, pixels (Paget and Longstaff, 1995) or pixel-regions (Cross and Jain, 1983) are represented by nodes in the network. The pixel (regions) are connected with edges to other *near-by* pixel (regions). The entire texture can be modelled by a Gibbs random field, for which the probabilities depend on a set of parameters, say θ. The parameter values are representative of the properties of the texture. MRF models have been used for texture synthesis (Cross and Jain, 1983), classification (Chellappa and Chatterjee, 1985) and segmentation tasks (Panjwani and Healey, 1995, Won and Derin, 1992).

Mandelbrot (1983) argues that a large proportion of natural scenes can be described using *fractals*. Fractals are defined as a self-similarity group. They are formed by the union of an object and scaled-down copies of the object. The scale ratio, r, is kept constant for successive down-scaling steps (Mandelbrot, 1983). Early attempts at texture classification based on fractal methods relied on properties of the *fractal dimension*. The fractal dimension, given by the ratio of the log of repeat elements to the log of $\frac{1}{r}$, is a measure of a fractal's roughness/complexity (Pentland, 1984, Sarkar and Chaudhuri, 1992). Mandelbrot also explains that two very different patterns can have very similar fractal dimensions. He introduces the concept of *lacunarity*, which can be used as a measure of how many *gaps* a pattern has. Keller et al. (1989) show that in two simple segmentation tasks, a segmentation based on lacunarity performs better than fractal dimensions alone.

2.5.4 Structural/Geometric Methods

Structural methods are based on the assumption that textures are made from repeating *entities*, called *texels*, which constitute the *microtexture*. The *macrostructure* is given by the set of placement rules that governs the arrangement of these entities (Haralick, 1979). Structural texture analysis methods are often carried out in two key stages. The first is the identification of the *microtexture*. The second step is to infer the set of rules governing the arrangement of the *microtexture*. Because structural methods operate under the assumption that there is a repeating unit, the performance of these models is often only useful for fairly-regular textures (Tuceryan and Jain, 1998).

A fairly recent structural/model-based approach is epitomic texture analysis. The epitome of a textured region is defined as a *mini-version* of the texture, or a condensed version of the texture. In epitomic analysis, a texture, or more generally an image, can be defined by its epitome and a function mapping the epitome to the image pixel values (Jojic et al., 2003). Epitomic analysis is very useful in image compression. This is especially so for video compression where regions in consecutive frames in a video are likely to be similar. These regions, repeating from frame to frame, can be represented by a single epitome.

2.5.5 Machine Learning Methods

Since high power computers have become available to researchers, an increasing number of machine learning methods have been developed in image classification, texture recognition and segmentation (Cimpoi et al., 2016, 2015), and medical image analysis (Shen et al., 2017) in applications for facial recognition (Sun et al., 2014), X-ray image analysis (Rogers et al., 2017), pedestrian detection (Ouyang and Wang, 2013), and even kinship verification (Boutellaa et al., 2017). While these methods have achieved performance levels far superior to more 'hand-crafted' methods (LeCun et al., 2015), edge cases are still not well understood (Andrews et al., 2016, Tanay and Griffin, 2016). In this work we chose to err on the side of predictability and intuition over performance so we do not explore these methods in learning. Not withstanding this, we acknowledge their feats in reaching the performance levels they have.

2.6 Relevant Methods for Texture Scale

As discussed in Section 2.1, many of methods for estimating the scale of texture in the literature aim to identify the repeating element (texel) within the image. The size of the texel or distance between neighbouring texels is taken as a measure of scale. These methods are ill-defined for irregular/stochastic textures which are not formed by the placement of a repeating texel on the scene, and for homogeneous textures, in which the texel is not placed on the scene according to a regular grid. The exceptions to this are the methods of Ardizzone et al. (2013), which is designed

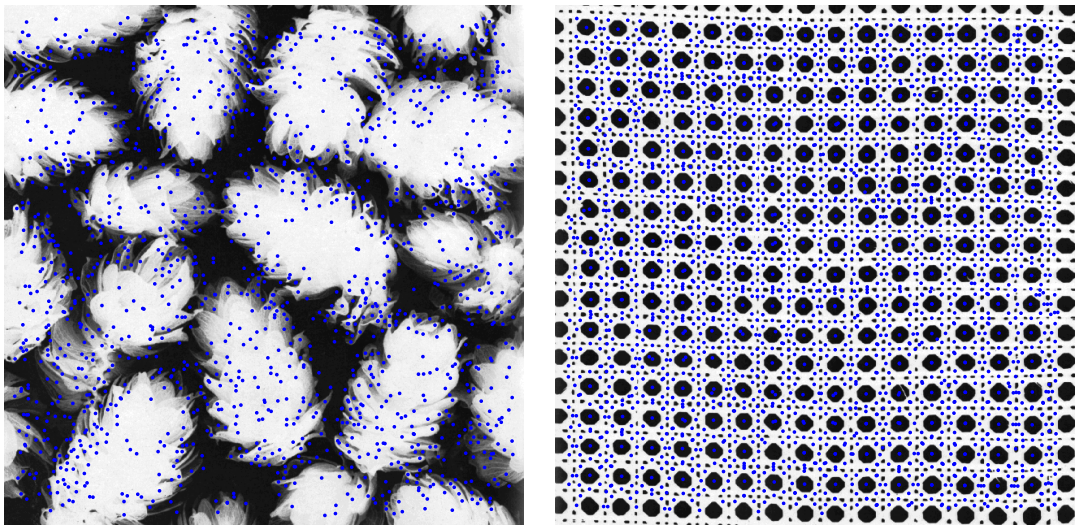


Figure 2.10: Surf keypoints for Brodatz D88 (left) and D101 (right).

to work for homogeneous as well as for regular and near-regular textures, and that of Hong et al. (2008) which is designed to work for textures of all regularities. We present these two methods in detail in this section because we will compare the accuracy of the estimates obtained using these two algorithms, to that of those obtained using our own algorithm, which we present in Chapter 4.

Ardizzone et al. (2013) propose a method for estimating the global scale of textures. They define texture scale as the size of the repeating unit forming the texture, or texel. The algorithm they propose is designed to work for textures displaying constant texture, i.e. containing a single type of texture, and formed by a repeating unit, i.e. for regular, near-regular and homogeneous textures. The Brodatz images are known for displaying constant texture, so the authors use them in the development of their tool.

The method comprises four steps:

1. Compute the image keypoints
2. Calculate 'keypoint density maps' for a range of window sizes
3. Plot the square root of the modal number of keypoints in the keypoint map *vs.* window size
4. Identify the subset of window sizes for which the modal number of keypoints

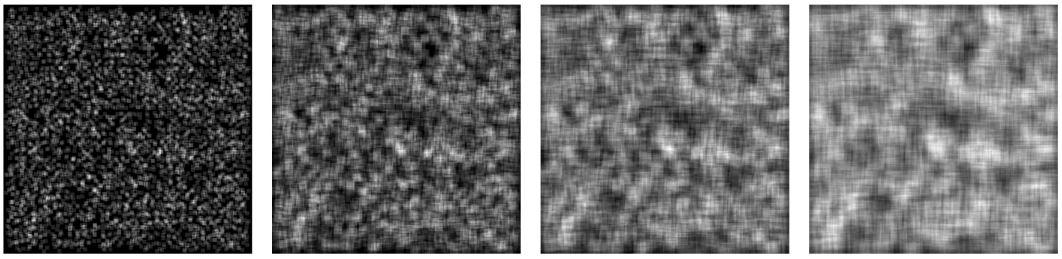


Figure 2.11: Keypoint density maps for radii of 11, 21, 31 and 41 pixels respectively on Brodatz 88.

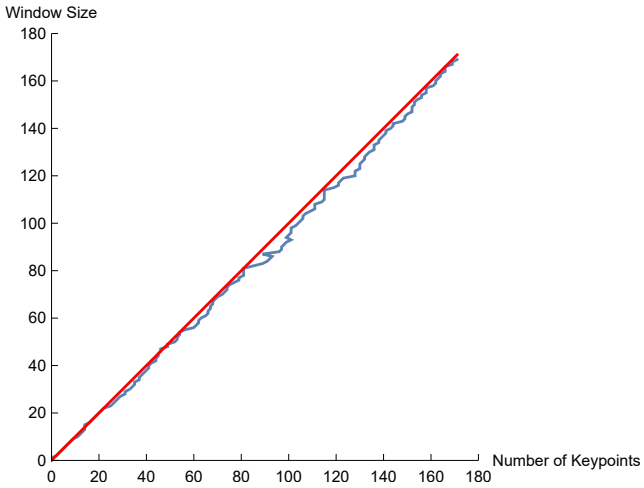


Figure 2.12: Normalised number of keypoints *vs.* window size for Brodatz D88.

is closest to linear

Figure 2.10 shows the SURF keypoints for two Brodatz textures. For each pixel in the image, and for a given window size, we can calculate the number of points fitting in the windows centred at the pixel. Figure 2.11 shows the keypoints maps for Brodatz D88 calculated using upright square windows of 11, 21, 31 and 41 pixels in height respectively, left to right.

For an image of $R \times C$ pixels, with p keypoints, the mean number of pixels per keypoint, $\bar{P}$ is:

$$\bar{P} = \frac{R \cdot C}{p} \tag{2.7}$$

Therefore, by construction, expected number of keypoints in an upright square window of size $S_x \times S_x$ is x, where:

$$S_x = \sqrt{x \cdot \bar{P}} \tag{2.8}$$

In practice the keypoints will not be uniformly distributed on the image, so Ardizzone et al. (2013) calculate the modal number of keypoints, $M(S_x)$ in a window of size S_x, for a range of window sizes. They then argue that the scale of a texture can be read as the step-size $\mathscr{S}$ which minimises the distance between the observed modal number of keypoints, and the window size. To identify the step-size, they sub-sample $M(S_x)$ at a range of step size; $\{x_i\}_i$, so that $M_n(x_i) = M(n \cdot x_i)$. The texture scale, $\mathscr{S}$, is thus given by:

$$\mathscr{S} = \arg\min_{x_i} \sum_i (M_n(x_i) - n \cdot x_i)^2 \tag{2.9}$$

Hong et al. (2008) propose a method for estimating the scale of texture at each pixel of the image based on energy minimisation. One of the ways in which the authors demonstrate the efficacy of their method is by showing that the local scale estimates obtained on Brodatz textures are approximately constant across each Brodatz. This is expected as the Brodatz exhibit constant texture, i.e. they only show one type of texture. This is very relevant for us, as we take the mean of the local scales for each Brodatz obtained using this method as the global scale estimate for the texture when comparing the performance of this algorithm to ours in Chapter 4. In their work, they showcase how their scale map can be used for segmentation tasks on natural images.

For a given point $p = (p_1, p_2) \in \Omega$ in image $I : \Omega \subset \mathbb{R}^2 \to [0,1]$, the window $\mathscr{W}_{p,r}$ centred at p and with 'radius' r is defined as:

$$\mathscr{W}_{p,r} = \left\{ q = (q_1, q_2) \in \Omega : \max_{1,2} \{|p_1 - q_1|, |p_2 - q_2|\} < r \right\} \tag{2.10}$$

The eight neighbouring windows to this $\mathscr{W}_{p,r} = \left\{\mathscr{W}^N_{p_i,r}\right\}_{i=1}^8$ is defined as follows:

$$\begin{aligned} \mathscr{W}^N_{p,r} = \{ & \mathscr{W}_{(p_1+2r,p_2+2r),r}, \mathscr{W}_{(p_1,p_2+2r),r}, \mathscr{W}_{(p_1-2r,p_2+2r),r}, \mathscr{W}_{(p_1-2r,p_2),r}, \\ & \mathscr{W}_{(p_1-2r,p_2-2r),r}, \mathscr{W}_{(p_1,p_2-2r),r}, \mathscr{W}_{(p_1+2r,p_2-2r),r}, \mathscr{W}_{(p_1+2r,p_2),r} \} \end{aligned} \tag{2.11}$$

Figure 2.13 illustrates this visually for $r = 1$ (right) and $r = 3$ (left). The central

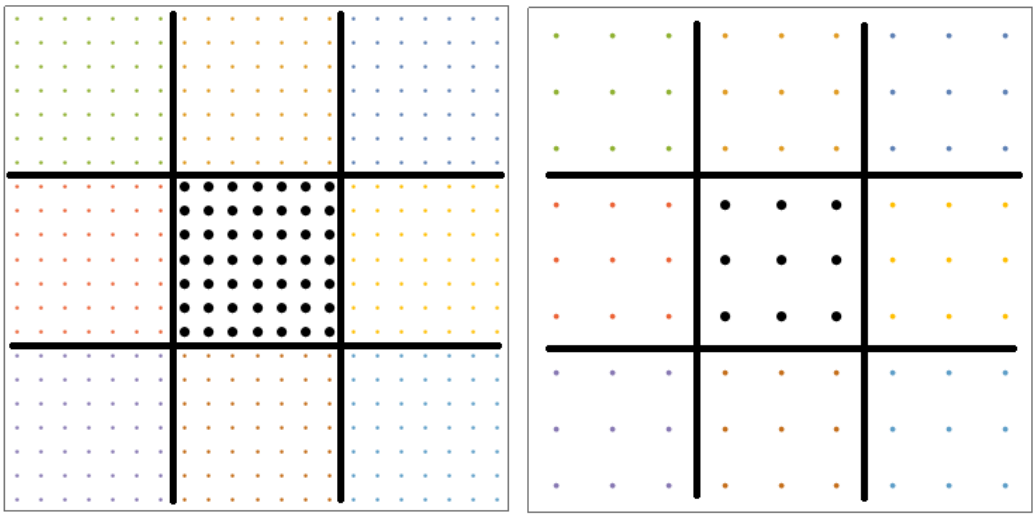


Figure 2.13: Hong et al. (2008) compare the energy between adjacent tiled regions of different sizes. Each dot corresponds to a pixel. Black region is $\mathscr{W}_{p,r}$, mid points of back region is p, and radii are 3 and 1 for left and right respectively. Coloured arrays correspond to each of the 8 $\mathscr{W}^N_{p,r}$.

black pixel in each of the two arrays corresponds to p, the black dots correspond to $\mathscr{W}_{p,r}$ and the 8 coloured areas correspond to each of the eight neighbouring windows in $\mathscr{W}^N_{p,r}$.

The probability density function for each of the windows $P_{\mathscr{W}_{p,r}}$ is defined by the gray-scale histogram of intensity values from $\mathscr{W}_{p,r}$. An energy function $\mathscr{E}$ is defined as follows:

$$\mathscr{E}_{p,r} = D\left(P_{\mathscr{W}_{p,r}}, P_{\mathscr{W}^N_{p,r}}\right) - \alpha H\left(P_{\mathscr{W}_{p,r}}\right) + \beta r \tag{2.12}$$

Where $D(P_{\mathscr{W}_1}, P_{\mathscr{W}_2})$ is a dissimilarity measure (distance) between $P_{\mathscr{W}_1}$ and $P_{\mathscr{W}_2}$, and $H(P_{\mathscr{W}})$ of the distribution, $P_{\mathscr{W}}$, of window $\mathscr{W}$. The authors consider the Kullback-Leibler divergence, and the Wasserstein distance as candidate measures of dissimilarity. They opt to use the latter as the Kullback-Leibler divergence is undefined when $P_{\mathscr{W}_2}$ is zero, which occurs frequently with gray-scale histograms of small regions. Figure 2.14 shows a synthetic texture (left) and the KL divergence (red, right), Wasserstein distance (green, right) and entropy (blue right), of the synthetic texture on the left side of the image, with varying window size. The window size w, is defined as $w = 2r+1$. The values of design parameters α and β are tuned

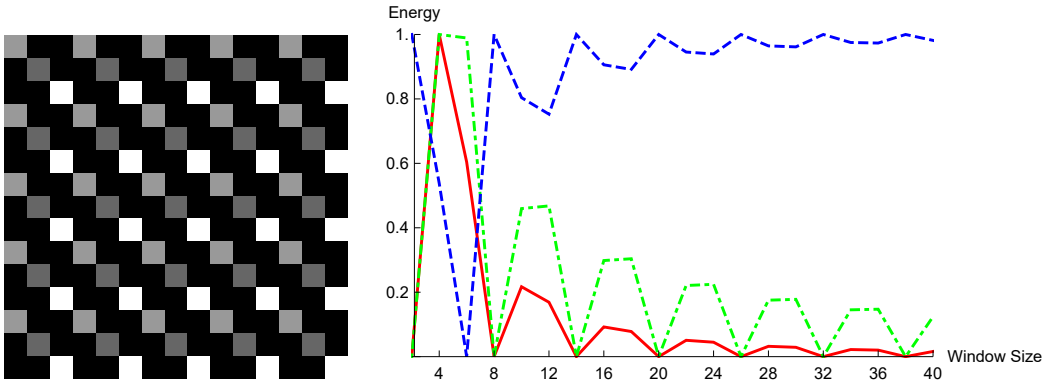


Figure 2.14: Texture scale method by Hong et al. (2008). Three energy function components: KL divergence (red, right), Wasserstein distance (green, right) and entropy (blue right) for a synthetically generated texture (left).

on the Brodatz and fixed as design parameters with values $\alpha = 0.001$, and $\beta = 0.1$.

Figure 2.15 shows the Wasserstein distance (top left), KL divergence (top centre) and entropy for 10 points on Brodatz D101 (bottom left). The curves are combined as per Equation 2.12, and the texture scale at p is given by:

$$S = \inf_{r} \min_{i} D_W \left(P_{\mathscr{W}_{p,r}}, P_{\mathscr{W}^N_{p,r}} \right) - \alpha H \left(P_{\mathscr{W}_{p,r}} \right) + \beta r \tag{2.13}$$

The firs component of the energy function that we are minimising finds a size that minimises the distance between the histograms of neighbouring reasons. The second component, with the entropy, tmaximises the complexity of the window in order to avoid picking up constant regions within a texel. The third component penalises for large windows so that the we avoid picking regions containing multiple texels - these will generally be captured by subsequent local minima.

The bottom right of Figure 2.15 shows a texture scale heat-map for Brodatz D101 (figure 2.15, bottom left). As expected by the authors, the scale is fairly constant. We take the mean of the texture scale values across the image as a measure of global scale on the Brodatz.

2.7 Summary from Literature Review

In this chapter we explored the relationship between texture scale, regularity and periodicity, provided a brief introduction to measurement theory in the context of

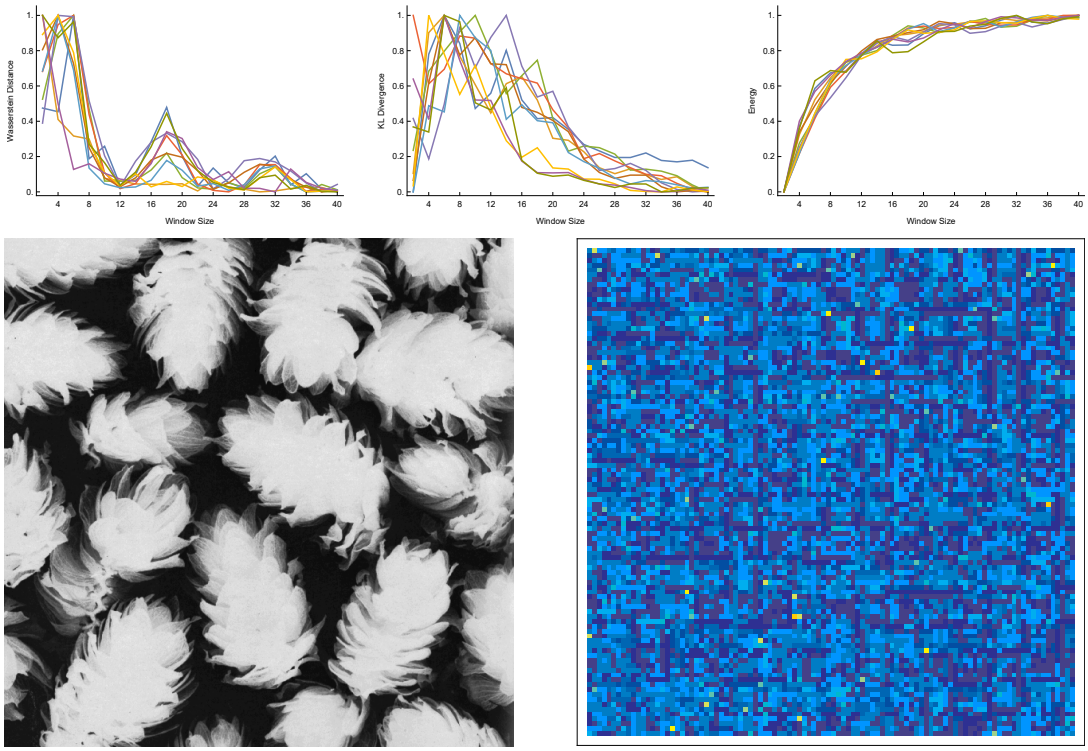


Figure 2.15: The three components of the energy curve, for Brodatz D101 and the corresponding scale map according to Hong et al. (2008).

perceptual studies and introduced one of the most common psychophysics scaling methods to set the scene around measuring human assessment of texture scale, the first part of the research covered in this work (Chapter 3). We then introduced some texture databases and applications in which they have been used and a brief taxonomy on texture methods. Finally, in Section 2.6, we reviewed in detail the two methods most similar to ours in scope. Next, we will introduce the two perceptual studies we carried out in order to capture human assessment of texture scale before presenting our empirical algorithm for the quantification of texture scale,

Chapter 3

Quantifying Human Perception of Texture Scale

In the last two chapters we explored the literature on texture scale and the different meanings of the term *scale* when used in the context of texture. In Chapter 2 we showed how the definition varies according to the application at hand depending on whether or not the texture is formed by a spatial arrangement of a discernible repeating unit.

In this chapter we show that humans can assess texture scale consistently, irrespective of texture regularity, and measure their assessment of texture scale, without making any claims on the physiological process behind it.

As previously mentioned, there are no well established definitions for texture scale when referring to non-regular textures. Attempts at characterising texture scale have been made by Ardizzone et al. (2013) and Hong et al. (2008). The characterisation by Ardizzone et al. (2013) relies on the spatial distribution of image features. Hong et al. (2008) use a pixel by pixel approach based on the similarity of histograms of luminance values in neighbouring regions.

Neither of these definitions were suitable for our characterisation as they were not perceptually validated and required the windows to be upright squares.

Instead we wanted to characterise texture scale perceptually, devoid of any underlying models for texture or physiological mechanisms. We considered several characterisations, but two of them are preferred. The first is “the length scale over

which texture appears consistently", and the second is "the window size beyond within which texture appears consistently".

Both contain the key idea of regions of consistent appearance. Although perceptually motivated, such a characterisation is highly consistent with various paradigms for texture (regions of constant statistics, a repeat element + a placement rule, ...), without being too restrictive as to how these ought to be interpreted. In the case of the definition of *area with constant statistics*, we do not specify what the word *constant* is supposed to mean. When thinking of it as a repeating element + placement rule, we do not impose the number of repeating elements that need to be considered in order to understand the placement rule.

Given the lack of a mathematically rigorous definition for texture scale, we have decided to propose a perceptual characterisation. When doing so, we are often asked "why should we care about the perception of a property, as opposed to the property itself?". While we understand the practical benefits of having a more restrictive definition of texture scale, in situations where the objectives are to both *(i)* quantify human perception (without introducing any form of experimental bias) and *(ii)* devise a scale that is communicable and easily interpreted by researchers across disciplines, it is often best to start with a perceptual characterisation.

A drawback of such an openly interpretable characterisation is that any measurements we make will be affected by *(i)* the true scales, *(ii)* population wide bias, and *(iii)* individual bias.

In this chapter we present two psychophysical experiments we have carried out in order to *(a)* validate our hypothesis that people are able to assess texture scale consistently, and *(b)* quantify their assessment with the aim of developing an algorithm for the automatic quantification of texture scale. The subjects were PhD students and academics from a range of scientific disciplines (vision, computer science, biology, engineering, mathematics).

Before we introduce the experiments, we justify our choice of the Brodatz dataset for this study in Section 3.1. There, we also explain how we have divided the dataset according to regularity, and isotropy for a more detailed analysis of both

the subjects' responses and the performance of the proposed algorithm.

The first experiment, introduced in Section 3.2 is a direct scaling experiment in which subjects were asked to make an absolute judgement on the scale of the textured image with which they were presented. The results from this experiment give us confidence that our hypothesis on the consistency of human assessment of texture scale, irrespective of texture regularity, is correct.

While the results from the first experiment provide validation for our hypothesis, there are two main drawbacks to such an approach in our case. First, in order to obtain sufficiently narrow confidence intervals using a dataset the size of the Brodatz, each observer would be required to give a huge number of responses (in the thousands). Second, when subjects are asked to make absolute judgements, their responses may be affected by individual bias.

The second experiment, introduced in Section 3.3, is a two alternative forced choice (2AFC) experiment which enabled us to derive an interval scale for perception, which has been shown to be useful (Protonotarios et al., 2014). However, this method is not problem-free either. In addition to possible issues with data quality, which we discuss later on, the main limitation of this method is that the derived texture scale values are on an internal-psychophysical interval scale. As a result, the derived values do not represent a physical "distance" or "area". This is particularly problematic in light of the fact that we want to obtain a measure of texture scale that can be easily communicated.

Nonetheless, it has been shown that the relationship between psychophysical interval scales and physical unit scales obey Weber's law (Fechner, 1966, Gescheider, 1997). Therefore, we can use the measurements in pixel units from the first experiment to convert the estimates arising for the 2AFC task, in jnd units to pixels. The details of this, as well as the resulting values for the Brodatz dataset, are presented in Section 3.4.

We conclude this chapter, in Section 3.5, with a discussion of the limitations of using perceptual studies to derive ground-truth datasets, possible extensions to the studies, and a more detailed analysis of the factors which affect subjects' judge-

ments.

3.1 Choosing a Suitable Dataset

In Section 2.4 we provided a detailed review of some of the most common texture datasets found in the image analysis literature. Having compared these candidate datasets across three dimensions (constancy within textures, variety between textures and dataset size), we chose to use the Brodatz dataset in our two perceptual studies.

In order to measure a subject's assessment of scale accurately, we wanted a dataset that would enable the subject to isolate texture from other aspects of the image. Therefore we looked for datasets with constant texture within images, no backgrounds, perspective, or other image features. The Brodatz dataset satisfies these requirements.

We wanted to verify subjects' ability to assess the scale of a wide range of textures. The natural image and material datasets (e.g. Brodatz) are a lot more varied than specialised datasets (e.g. tissue cell samples). While there are known biases affecting subjective assessment of image properties when presented with familiar objects (e.g. those found in natural images and material images), we did not believe that such effect would be pronounced for the Brodatz due to their constancy within textures.

In order to assess consistency within and between subjects, we needed a dataset that was not only highly varied but also small enough for any given subject to provide responses more than once for each texture. The 112 strong Brodatz is an appropriate size given the requirements.

Although we could have combined images from various datasets instead of picking one of many, we did not do this for two reasons. First, it could lead to questions with regards to the images included (e.g. were images selected *a posteriori* to optimise agreement with our hypothesis?). Second, we wanted our results to be useful for further research, and reporting results for a complete and well established dataset was more likely to achieve this.

We use a version of the Brodatz textures which is freely available for download and which has lightness inverted values. These images have a slight advantage for psychophysical experiments as the imaged textures are harder to recognise so may reduce the impact of semantic associations. We are happy to use a gray-scale database because there is evidence that colour does not have significant influence on the perception of texture complexity (Ciocca et al., 2015a,b).

The Brodatz dataset contains images with a range of regularities and isotropy levels. We rank the Brodatz set according to these dimensions and report the levels in Table 3.1. Textures are ranked 1-3 according to their regularity, where a score of 1 indicates irregular textures, and a score of 3 indicates regular textures. Textures are ranked 1-5 according to their isotropy, where a score of 1 indicates textures that are highly anisotropic and a score of 5 indicates textures that are highly isotropic.

3.2 Direct Scaling Approach

The first psychophysical experiment that we performed was a direct scaling one. This means that subjects were asked to make an absolute assessment, in this case about texture scale. The purpose of this first experiment was to provide us with preliminary validation that there is consensus on the assessment of the scale of textures, before we attempted automatic methods to estimate it, and validate those against data from the second experiment.

3.2.1 Method

Subjects were presented with a single texture and asked to drag-and-select a "window of the smallest size within which the texture appears consistently". The interface restricted subjects to choosing upright square windows. Subjects were restricted to dragging-and-selecting a square window in this study so that they were forced to reduce their judgement to a single figure. An experimental run involved a subject being presented with each of the 112 Brodatz textures in turn. Figure 3.1 shows a screen-shot of the interface. This was a web-based interface. Each subject was asked to perform the experiment on their own machines over the internet. We recorded the position and size of the selected window and the time it took subjects

Tex	Reg	Iso	Tex	Reg	Iso	Tex	Reg	Iso	Tex	Reg	Iso
D1	1	4	**D29**	3	4	**D57**	3	4	**D85**	1	1
D2	3	4	**D30**	3	4	**D58**	3	1	**D86**	3	3
D3	2	4	**D31**	3	4	**D59**	3	1	**D87**	2	2
D4	3	5	**D32**	3	5	**D60**	3	3	**D88**	2	3
D5	3	5	**D33**	3	5	**D61**	3	2	**D89**	2	2
D6	1	4	**D34**	1	4	**D62**	3	1	**D90**	3	2
D7	3	5	**D35**	2	3	**D63**	3	3	**D91**	3	1
D8	2	2	**D36**	2	3	**D64**	1	4	**D92**	3	5
D9	3	2	**D37**	3	1	**D65**	1	3	**D93**	3	1
D10	3	1	**D38**	3	1	**D66**	2	5	**D94**	1	4
D11	1	2	**D39**	3	1	**D67**	2	1	**D95**	1	4
D12	3	3	**D40**	3	1	**D68**	2	1	**D96**	2	1
D13	3	2	**D41**	2	1	**D69**	3	1	**D97**	3	2
D14	1	3	**D42**	2	1	**D70**	3	1	**D98**	3	4
D15	3	1	**D43**	3	1	**D71**	3	1	**D99**	3	2
D16	1	4	**D44**	3	1	**D72**	3	1	**D100**	3	3
D17	2	2	**D45**	3	2	**D73**	3	3	**D101**	1	5
D18	1	3	**D46**	1	1	**D74**	2	4	**D102**	1	5
D19	2	4	**D47**	1	1	**D75**	2	5	**D103**	2	5
D20	1	5	**D48**	1	4	**D76**	2	3	**D104**	2	5
D21	1	5	**D49**	1	1	**D77**	2	5	**D105**	2	2
D22	2	4	**D50**	3	1	**D78**	2	2	**D106**	2	2
D23	3	4	**D51**	3	1	**D79**	2	2	**D107**	3	4
D24	3	3	**D52**	1	4	**D80**	2	4	**D108**	3	4
D25	1	2	**D53**	1	4	**D81**	2	3	**D109**	3	4
D26	1	2	**D54**	3	5	**D82**	1	5	**D110**	3	4
D27	3	5	**D55**	1	3	**D83**	1	1	**D111**	2	5
D28	3	5	**D56**	1	2	**D84**	2	5	**D112**	2	3

Table 3.1: Brodatz textures scored according to regularity and isotropy. Regularity: 1 for irregular, 3 for irregular. Isotropy: 1 for anisotropic, 5 for isotropic.

to perform each trial. There was no time limit for the experiment. The order and orientation (in multiples of 90 degrees) of each image was randomised for each trial and between subjects. Subjects were presented with each of the 112 Brodatz in turn so that we could evaluate the consistency between experimental runs. In order to evaluate within-subject consistency, we allowed each subject a cooling period of 48 hours before asking them to perform the complete experiment a second time. While the experiment was performed in non-controlled conditions, we did not expect this to have a major impact on the results. The Brodatz images are 640x640 pixels, so

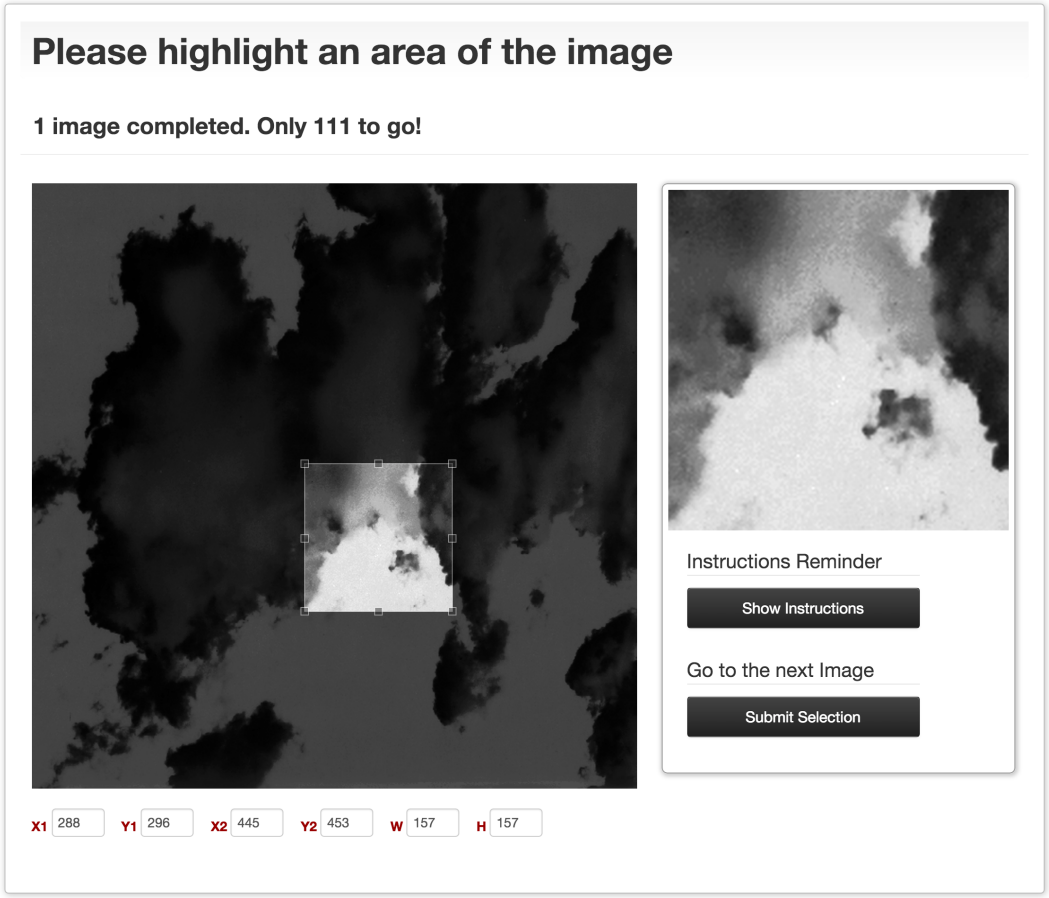


Figure 3.1: Interface used for the direct-scaling experiment. Subjects were asked to drag and select a window on the main interface pane (left). The interface restricted the subjects to selecting upright square windows. The selection was shown on the left pane of the interface. Users could then submit their selection or make another selection.

the likelihood of sub-sampling due to display resolution restrictions was very low. In addition, because we perform a Procrustes transformation (Gower and Dijksterhuis, 2004) to correct for individual bias when analysing the results (Section 3.2.2), much of the systematic inter-subject variability is removed.

3.2.2 Response Analysis

The histogram of selected window sizes (measured in pixels), shown in Figure 3.2 (left), is strongly positively skewed. To facilitate statistical analysis of this data we computed log window sizes, which resulted in a less skewed (more Gaussian looking) histogram of values (Figure 3.2, right). Henceforth we will use $\log_2$ window size, but refer to it simply as window size.

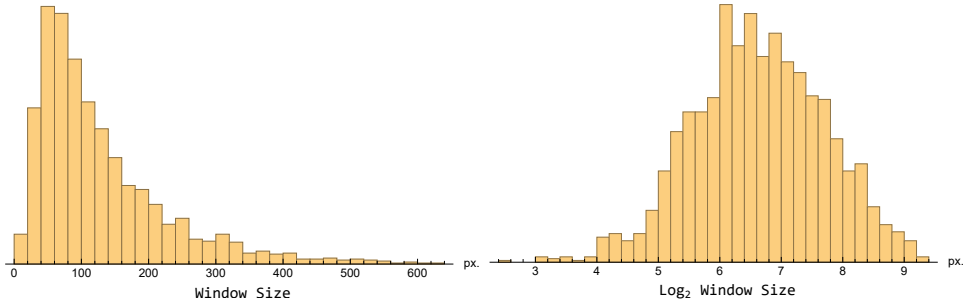


Figure 3.2: Histogram of subjects' responses to the direct-scaling experiment in pixels (left) and $\log_2$ pixels (right).

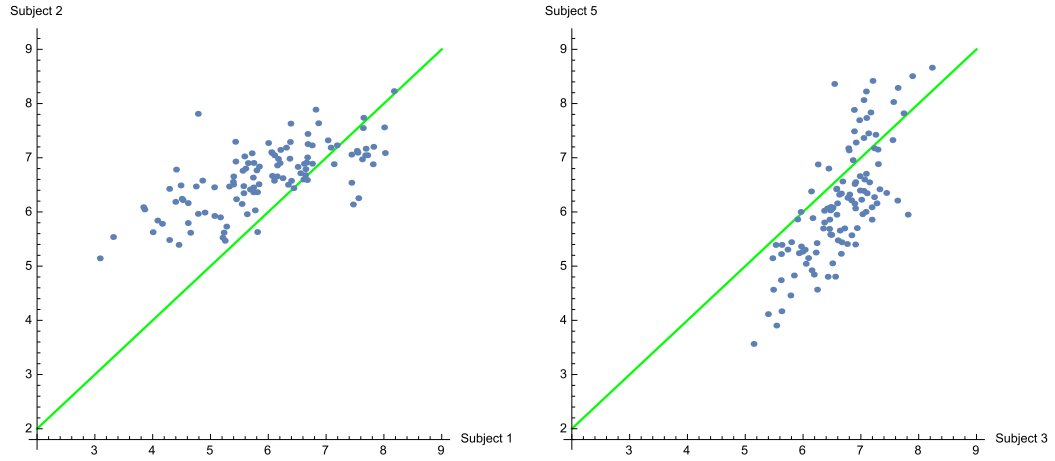


Figure 3.3: Scatter plots between two distinct subject-pairs illustrate systematic differences between subjects' window selections, and $y = x$ line (green).

We calculated the intra-subject Person's correlation for each subject's responses. These were all in the range 0.74-0.89, except for one subject for which it was 0.54. Given that this subject responded much faster than all the other subjects, the data was considered of poor quality and was discarded from all further analysis. The mean intra-subject correlation was then 0.83. This suggested that each subject had a consistent internal criterion for their selections.

We then averaged the two responses of each subject, and computed correlations between each pair of subjects. The mean inter-subject correlation was 0.73. The small difference (0.83 vs. 0.73) between the inter- and intra-subject correlations, indicated a substantial consensus across subjects' assessment of texture scale.

Further analysis of the responses revealed that the mean (over the 112 textures) of the standard deviation (over the eight subjects) of the log-window settings was

0.76, hence subjects' settings were varying $\pm 69\%$ around the mean setting (expressed in pixel units). This seemed rather large so we examined the data further to determine the cause. Figure 3.3 shows a scatter plot of the responses of two example subjects. The left scatter plot shows that subject 1 set small windows smaller, and large windows larger, when compared to subject 2. The correlation of the responses between the two subjects is 0.69 for the the LHS graph, and 0.88 for the RHS graph on Figure 3.3. Scatter plots for other subject pairs showed other, roughly linear, relationships between subjects' responses.

To deal with the subjects' aversion to overly small and large window sizes, we perform a procrustes transformation (Gower and Dijksterhuis, 2004); i.e. we linearly transformed each subject's data to agree maximally with the mean settings. Post-alignment, the mean standard deviation of log-responses decreased to 0.34, i.e. a variation of $\pm 27\%$ around the mean. Thus, the variation across subjects' responses is due to a combination of *(a)* variation in what subjects consider the best window size for a texture, and *(b)* variation in subjects' aversion to very large and very small window sizes. Our alignment process is intended to remove much of variation source *(b)*. Using both the raw responses, and the aligned dataset, we calculate a mean window size and standard deviation for each texture, and display some examples in Figure 3.4. For textures of high consensus across subjects with regard to the best window size, the mean-alignment had little effect on the window size (Figure 3.4, first column). For textures without a clearly discernible repeating element, the effect of the variable aversion to very small or very large windows was more significant. For these textures (Figure 3.4, second column) the effect of alignment was pronounced.

The experimental set-up enabled us to calculate the variances of the perceptual scale estimates for individual textures (across subjects). We found that the standard deviation for all responses (after alignment) was 0.66 $\log_2$ pixels. The mean per-texture standard deviation (after alignments) was 0.34 (see Figure 3.5). The ratio of the mean per-texture standard deviation to the total standard deviation was 0.52 ($= \frac{0.34}{0.66}$). This tells us how much signal we are capturing with this experiment.

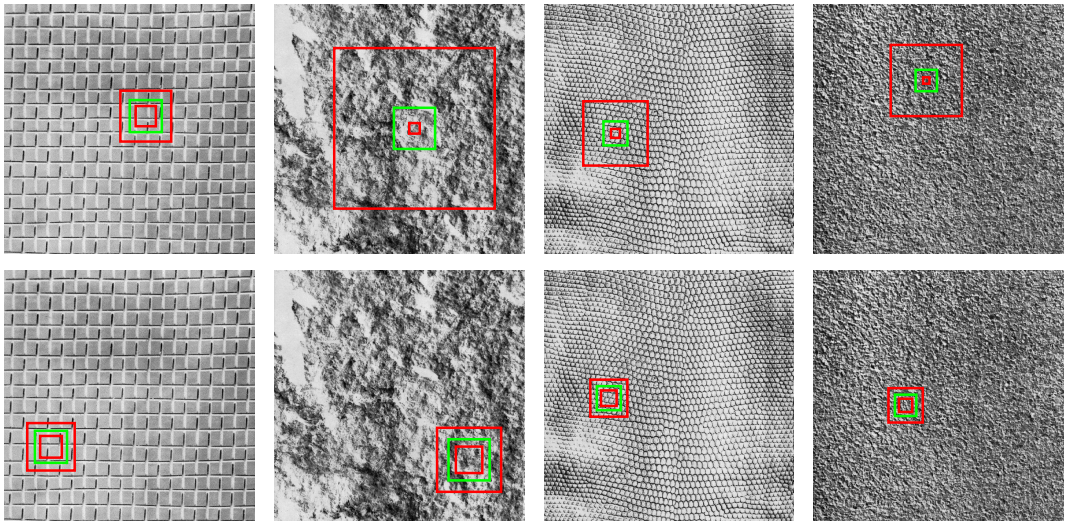


Figure 3.4: Mean (green) and mean ± 95% CI (red) of subject responses to the direct-scaling experiment (top row), and mean-aligned responses (bottom row) for four Brodatz textures.

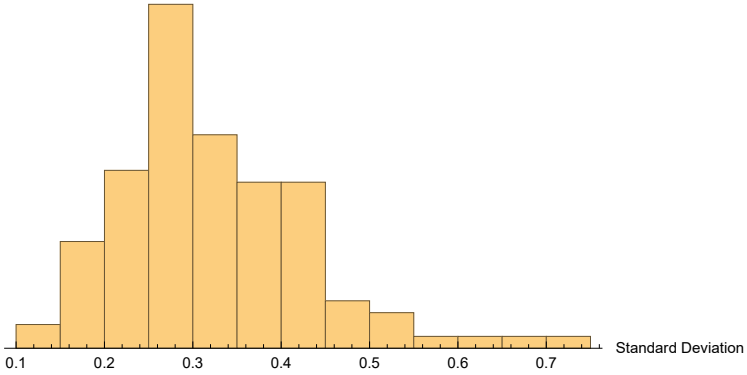


Figure 3.5: Distribution of the standard deviations of the mean-corrected subject responses to the direct-scaling experiment for the 112 Brodatz textures.

Figure 3.6 shows the four Brodatz textures for which the recorded responses had the largest associated standard deviations after alignment. We note that three out of the four are not homogeneous and one is highly anisotropic. These features explain the larger discrepancy between subjects' responses leading to the wider confidence intervals for the mean settings.

Subjects took on average 8.7 seconds to select a window size (standard deviation 4.4s). Figure 3.7 shows the distribution of trial durations.

Respondents prefer to select windows from the centre of the images. Figure 3.8 shows the selections of four different subjects. For each response by each subject

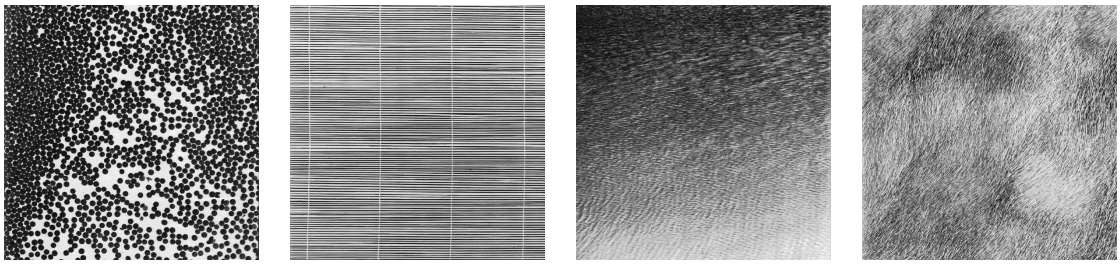


Figure 3.6: The four Brodatz textures (D76, D49, D38, D93) for which the recorded responses had the largest associated standard deviations after alignment in the direct-scaling experiment.

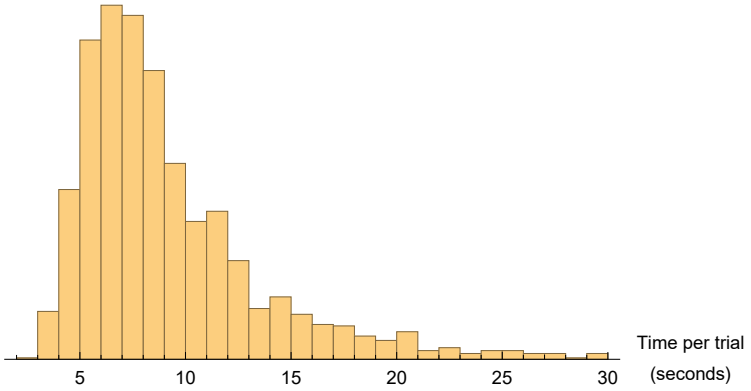


Figure 3.7: Distribution of number of seconds taken per response by subjects for the direct-scaling experiment.

we have highlighted the area selected for each of the 112 Brodatz textures.

We used our labelling of textures according to isotropy and regularity to investigate how subjects' responses varied across these dimensions. Figure 3.9 shows a scatter plot of the average mean log window size and standard deviation for the 112 Brodatz coloured according to isotropy. We notice that most of the black points are in the lower half of the plot while most of the red points are in the upper half. This

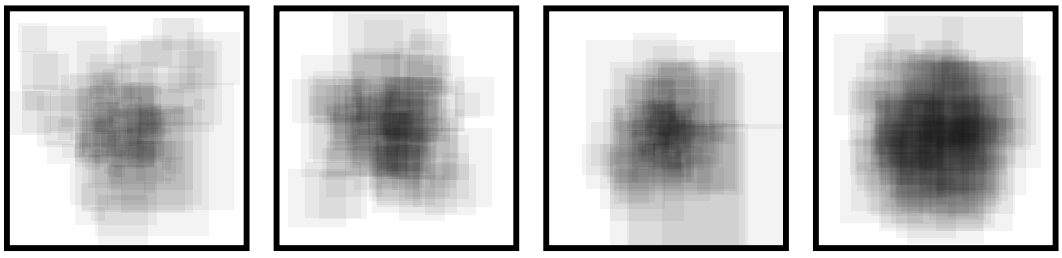


Figure 3.8: Each square shows the selections for each of the 112 Brodatz textures for four different subjects.

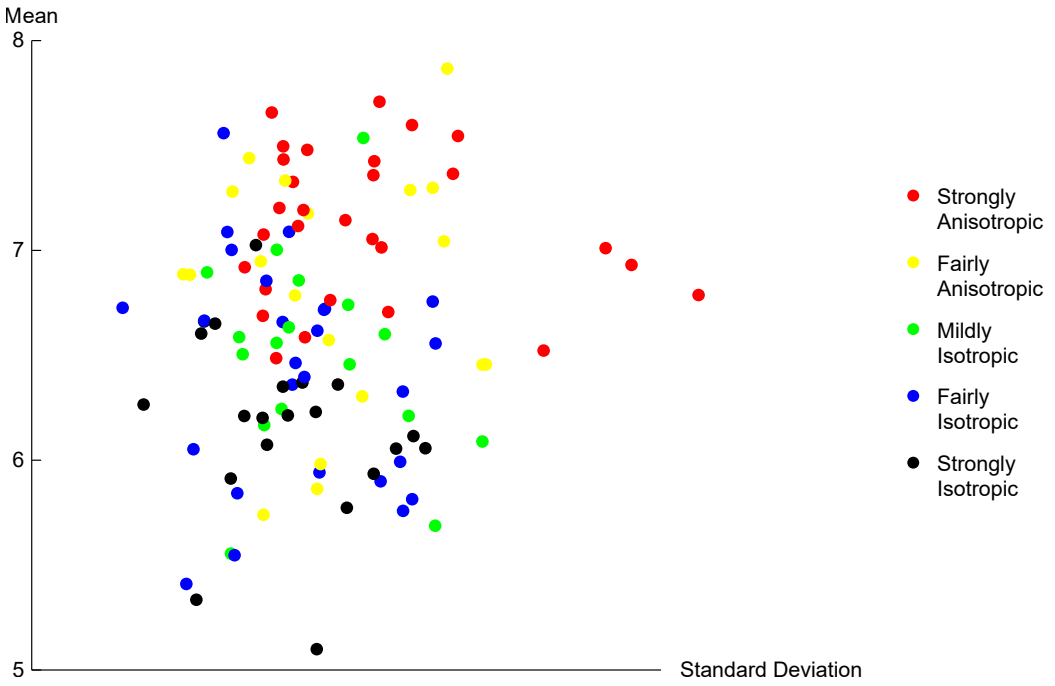


Figure 3.9: Variation of responses to the direct-scaling experiments according to isotropy.

indicates that subjects selected on average larger windows for the more anisotropic textures. Given the square-window restriction imposed by the interface. This suggests that subjects made larger, conservative selections, enabling them to see variation across the longer dimension of repetition.

Figure 3.10 shows a drop in the standard deviation of subjects' responses as textures become increasingly anisotropic (as per isotropy scores in Table 3.1). The standard deviation for the strongly anisotropic textures is on average 34% larger than that for the strongly isotropic ones.

3.2.3 Summary

Table 3.2 contains the linearly corrected perceptual based scales from the first, direct-scaling, perceptual experiment in $\log_2$ pixels and their corresponding standard deviations. Table 3.3 shows the same values in pixel units for ease of interpretation.

Through this experiment we have shown consistency in the assessment of texture scale across and within human observers. While these results provided us with preliminary validation for our hypothesis, they are not free of drawbacks. One draw-

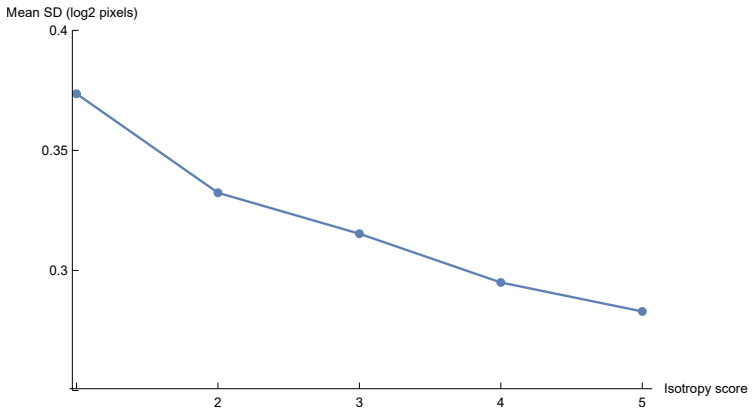


Figure 3.10: The responses for isotropic textures have shown a lower standard deviation compared to anisotropic textures in the direct-scaling experiment.

back is that subjects were restricted to selecting square windows which could affect the results for highly anisotropic textures. A second drawback is that, similarly to other direct scaling tasks, responses will reflect subjects' individual preferences (systematic bias). We performed a procrustes correction to account for this.

Our ambition was to devise an algorithm for estimating the scale of visual texture. To do so we needed a ground-truth dataset that could be used for the training and validation of our candidate algorithm. While the texture scale estimates derived from this first algorithm are more than good enough for the purpose of validating our initial hypothesis and motivating further work, the aforementioned limitations of this approach, coupled with relatively large standard deviations in relation to the estimates, do not make for an ideal ground-truth dataset.

In the next section (3.3), we present a second perceptual study devoid of these limitations which we later used as ground-truth values in the development of our algorithm.

3.3 Indirect Scaling Approach

In the previous section we presented the results from the first study (direct scaling). The results showed that there is a consistency between and within subjects regarding their assessment of texture scale for textures of all regularities.

The first study had two limitations, *(i)* the method is sensitive to systematic personal bias, and *(ii)* the square window restriction imposed by the interface. Even

Tex	Scale	SD	Tex	Scale	SD	Tex	Scale	SD	Tex	Scale	SD
D1	6.36	0.29	**D29**	5.99	0.41	**D57**	5.82	0.42	**D85**	6.59	0.3
D2	6.72	0.33	**D30**	7.09	0.22	**D58**	7.14	0.35	**D86**	6.89	0.19
D3	5.94	0.32	**D31**	7.09	0.29	**D59**	7.43	0.28	**D87**	6.57	0.33
D4	5.78	0.35	**D32**	5.94	0.38	**D60**	6.86	0.3	**D88**	7.53	0.37
D5	6.6	0.19	**D33**	6.37	0.3	**D61**	7.17	0.31	**D89**	7.44	0.24
D6	5.41	0.17	**D34**	5.76	0.41	**D62**	7.12	0.3	**D90**	7.04	0.46
D7	7.02	0.25	**D35**	6.46	0.35	**D63**	7	0.27	**D91**	7.32	0.29
D8	6.88	0.18	**D36**	6.09	0.5	**D64**	6.66	0.28	**D92**	6.21	0.28
D9	5.98	0.32	**D37**	6.92	0.23	**D65**	6.25	0.28	**D93**	6.79	0.74
D10	7.01	0.39	**D38**	6.93	0.67	**D66**	6.27	0.12	**D94**	6.76	0.45
D11	6.88	0.17	**D39**	7.36	0.47	**D67**	6.52	0.57	**D95**	6.56	0.45
D12	6.6	0.39	**D40**	7.42	0.38	**D68**	6.49	0.27	**D96**	7.59	0.42
D13	6.95	0.25	**D41**	7.48	0.31	**D69**	7.05	0.38	**D97**	7.3	0.44
D14	5.56	0.22	**D42**	7.49	0.28	**D70**	7.07	0.26	**D98**	6.73	0.1
D15	6.71	0.39	**D43**	7.65	0.27	**D71**	7.19	0.3	**D99**	7.33	0.28
D16	5.9	0.39	**D44**	7.7	0.39	**D72**	7.2	0.27	**D100**	6.56	0.27
D17	6.46	0.5	**D45**	7.28	0.22	**D73**	6.63	0.29	**D101**	6.06	0.4
D18	6.59	0.23	**D46**	7.54	0.47	**D74**	6.67	0.19	**D102**	6.08	0.26
D19	6.62	0.32	**D47**	7.36	0.38	**D75**	6.65	0.2	**D103**	6.06	0.44
D20	6.23	0.31	**D48**	7.56	0.21	**D76**	6.21	0.42	**D104**	6.12	0.42
D21	5.1	0.32	**D49**	7.01	0.64	**D77**	5.34	0.19	**D105**	6.46	0.5
D22	6.05	0.18	**D50**	6.76	0.33	**D78**	5.74	0.25	**D106**	6.31	0.37
D23	6.72	0.32	**D51**	6.81	0.26	**D79**	5.87	0.32	**D107**	6.85	0.26
D24	6.17	0.26	**D52**	5.85	0.23	**D80**	6.4	0.3	**D108**	7	0.22
D25	7.86	0.46	**D53**	5.55	0.22	**D81**	6.51	0.23	**D109**	6.46	0.29
D26	7.29	0.42	**D54**	6.21	0.24	**D82**	6.2	0.26	**D110**	6.33	0.41
D27	6.66	0.19	**D55**	5.69	0.44	**D83**	6.69	0.26	**D111**	6.36	0.34
D28	6.35	0.28	**D56**	6.79	0.29	**D84**	5.91	0.22	**D112**	6.74	0.35

Table 3.2: Brodatz texture scale estimates and their standard deviation in $\log_2$ pixels obtained from the direct-scaling experiment. Values shown are linearly-corrected to population mean in order to correct for systematic subject bias.

though we proposed a proscrutes correction, we chose to perform another experiment which is less sensitive to such variation which we present in this section.

Like in the previous study, this experiment was performed in non-controlled conditions,given the resolution of the Brodatz (640x640 pixels), and the fact that the judgements were relative, we expect little effect from sub-sampling environment-related inter-subject variability.

Tex	Scale	SD	Tex	Scale	SD	Tex	Scale	SD	Tex	Scale	SD
D1	98	18	**D29**	86	21	**D57**	81	20	**D85**	112	20
D2	129	28	**D30**	154	22	**D58**	171	44	**D86**	137	18
D3	83	17	**D31**	155	30	**D59**	204	36	**D87**	116	22
D4	79	17	**D32**	85	18	**D60**	136	27	**D88**	211	39
D5	116	13	**D33**	104	18	**D61**	170	30	**D89**	194	23
D6	64	14	**D34**	73	21	**D62**	159	29	**D90**	160	50
D7	153	27	**D35**	109	27	**D63**	149	31	**D91**	187	40
D8	137	17	**D36**	92	28	**D64**	116	20	**D92**	93	17
D9	88	12	**D37**	140	27	**D65**	92	16	**D93**	146	63
D10	148	36	**D38**	164	66	**D66**	94	9	**D94**	130	43
D11	134	16	**D39**	194	53	**D67**	119	44	**D95**	113	38
D12	122	31	**D40**	201	50	**D68**	110	16	**D96**	219	51
D13	145	23	**D41**	206	41	**D69**	157	43	**D97**	186	49
D14	68	12	**D42**	206	33	**D70**	158	35	**D98**	121	6
D15	126	29	**D43**	230	45	**D71**	170	43	**D99**	180	30
D16	83	16	**D44**	247	70	**D72**	172	35	**D100**	115	23
D17	109	28	**D45**	180	34	**D73**	120	20	**D101**	85	21
D18	112	15	**D46**	210	54	**D74**	118	11	**D102**	85	15
D19	119	26	**D47**	183	37	**D75**	117	10	**D103**	89	21
D20	91	19	**D48**	210	31	**D76**	97	32	**D104**	92	20
D21	60	13	**D49**	159	47	**D77**	64	9	**D105**	114	31
D22	85	10	**D50**	131	35	**D78**	75	14	**D106**	101	18
D23	123	21	**D51**	132	21	**D79**	82	15	**D107**	135	23
D24	92	12	**D52**	77	14	**D80**	103	17	**D108**	149	30
D25	268	69	**D53**	69	13	**D81**	110	17	**D109**	108	17
D26	177	41	**D54**	94	12	**D82**	91	14	**D110**	103	27
D27	118	18	**D55**	72	21	**D83**	119	19	**D111**	101	23
D28	101	19	**D56**	125	20	**D84**	81	7	**D112**	130	34

Table 3.3: Brodatz texture scale estimates and their standard deviation in pixels obtained from the direct-scaling experiment. Values shown are linearly-corrected to population mean in order to correct for systematic subject bias.

.

3.3.1 Method

This experiment is a 2AFC (two-alternative-forced-choice) type of experiment. This means that subjects were asked to make a choice between two alternatives, and were forced to do so (i.e. they were not able to skip a question, or select none or both alternatives). Before the experiment subjects were provided with our characterisation of texture scale (in writing) and then in turn presented with random pairs of Brodatz textures, randomly oriented in multiples of 90 degrees, and were asked to select the

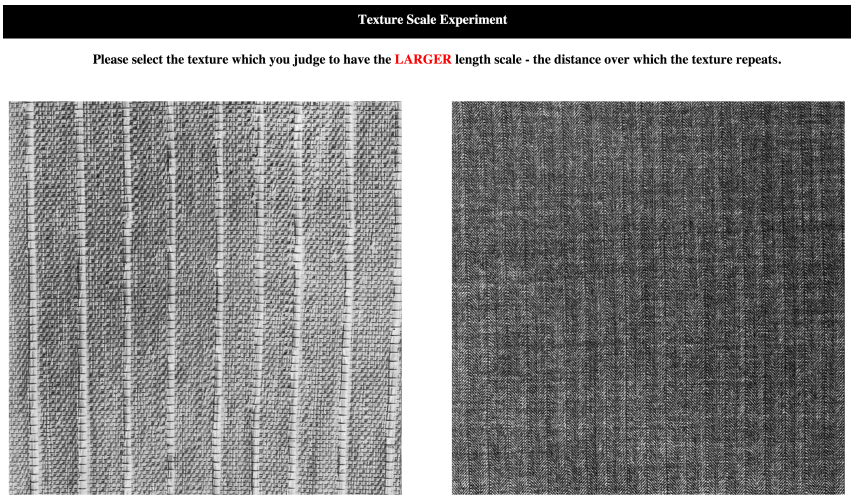


Figure 3.11: Interface for the indirect-scaling (2AFC) experiment.

one that they judged to have the larger texture scale. A screen-shot of the interface can be seen in Figure 3.11.

Subjects were asked to perform the experiment at home on their own machines. We asked subjects to perform at least 100 comparisons, and no more than 500, but did not exclude subjects who performed more than 500 or fewer than 100 comparisons. A total of 40 subjects participated in the study performing between 50 and 700 comparisons each. We collected a total of 13,950 judgements from forty subjects. This experiment was devoid of any window shape restrictions.

3.3.2 Building an Interval Scale

In the direct-scaling experiment we recorded subjects' absolute selections and were able to obtain texture scale estimates by taking the mean and standard deviations per subject. In this experiment, subjects are not asked to make absolute assessments, but to select one texture from a pair. We analysed these responses using an indirect scaling model (Protonotarios et al., 2014). The null hypothesis is that the noise/error in each realisation is normally distributed (Thurstone model, case V, Thurstone (1927)). The likelihood of the responses (e.g. texture D3 *vs.* D1) for each pair of textures is computed from a binomial distribution. Under this assumption, there exists a monotonic *preference* function, taking the form of a cumulative Gaussian. This monotonic preference function maps signed differences between the two

true-values to the probability that one will be preferred over the other. This model assumes that each perception of the scale of texture i is a random zero-mean Gaussian perturbation of a true attribute value, s_i. By assumption, the standard deviation of the random perturbations is the same for all textures. Hence, according to the model, a preference function $P(i,j) = \lambda + (1-\lambda)\,\mathrm{erf}\left(\frac{s_i - s_j}{0.477}\right)$ gives the probability that texture i will be judged as having a larger scale than the texture j. erf is the Gauss error function, and λ is the lapse rate, the probability that a subject makes an erroneous response (Wichmann and Hill, 2001). The constant 0.477 is chosen so that a scale difference of 1.0 units gives rise to a preference probability of 75%. By convention, such a distance is called a just-noticeable-difference (jnd) (Torgerson, 1958).

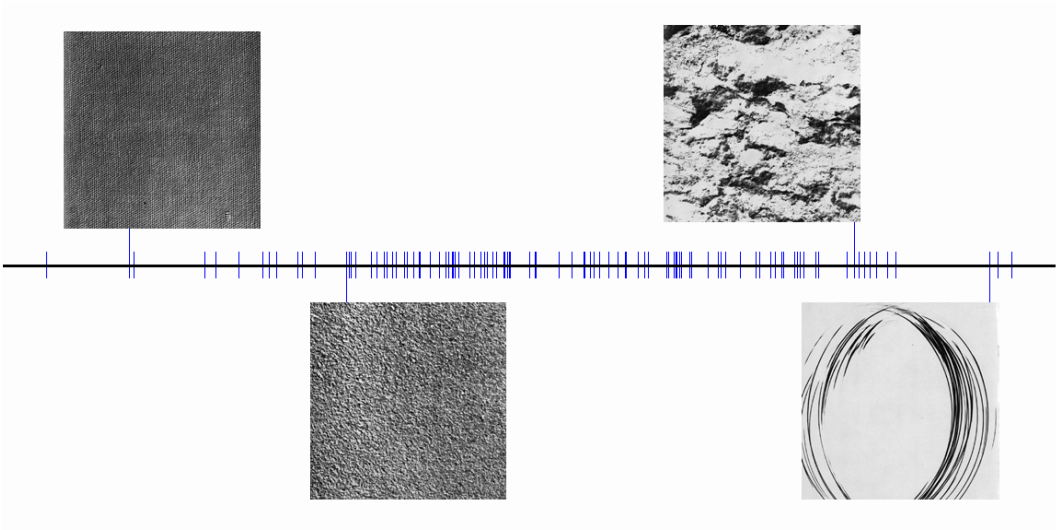


Figure 3.12: Four Brodatz textures on the derived interval scale from the indirect-scaling responses.

The arising scale from this method has an interval-scale structure. This means that equal distances at different parts of the scale correspond to equal discriminability. Interval-structure scales are invariant under multiplication and shift by a constant.

We used maximum likelihood (ML) fitting to estimate the texture scales and the lapse rate, as outlined in . We used gradient descent and multiple random starts to ensure that the optimum fit is found. Our fitting estimates a lapse rate of 6.4%

and a range of scale values spanning 11.7 jnds.

Insert formulas

An obstacle that we encountered when fitting the data was that our maximum likelihood approach was trying to assign overly large estimates to the four Brodatz textures with the largest associated estimates. These are shown in Figure 3.13. We looked at the raw data and found that there were 962 comparisons involving these four textures, only 8 of which involved a comparison between two of them. These 8 comparisons did not allow the ML maximiser to converge despite trying multiple start points and increasing the maximum number of steps which the maximiser was allowed perform. To address this, instead of trying to solve for the texture scale estimates, we solved for $20\sin X$, where X is the estimate. This trigonometric envelope function restricted the ML to assign values between -20 and 20 *jnds* to each texture. A perceptual attribute on an interval scale spanning Y units indicates that observers are able to distinguish Y levels of that attribute. We chose to restrict the range of values to $[-20, 20]$ because *(i)* it is unlikely that observers would be able to differentiate more than 40 levels of scale on the Brodatz (640x640 resolution), and *(ii)* increasing the width of the range of allowable values did not result in an increased range of the fitted values.

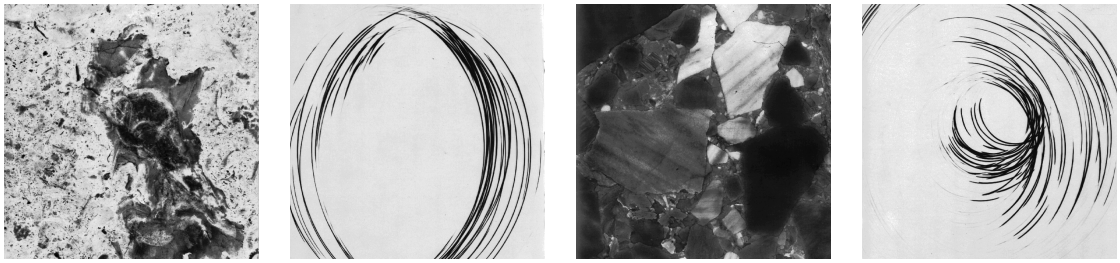


Figure 3.13: Four Brodatz textures with unusually large variances (D58, D44, D59, D43) in the responses to the 2AFC experiment. One common feature of these textures is that they are not formed from a discernible repeating element.

The goodness-of-fit of the model was assessed by comparing the empirical deviance to the distribution of deviances that result from Monte Carlo generated datasets from the ML model. Figure 3.14 shows the distribution of 10,000 Monte Carlo generated deviances. The empirical deviance from our model, 4246, falls

inside the 95% interval of acceptable deviances for 10,000 repetitions, $[4018, 4329]$, thus the derived interval-scale model is accepted.

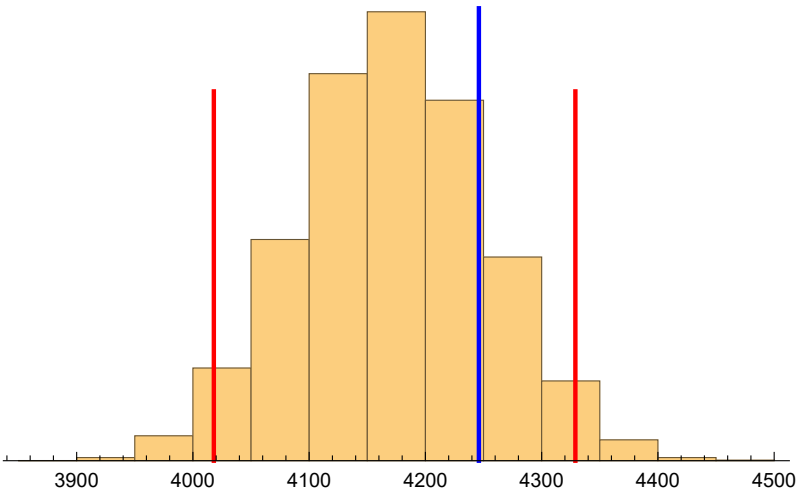


Figure 3.14: Histogram of deviances for the 10,000 Monte Carlo generated datasets. The empirical deviance (blue line) falls within the 95% confidence interval for the 10,000 generated dataset. This indicates that the derived interval scale from the 2AFC interval scale can be accepted as valid.

3.3.3 Response Analysis

Due to the nature of the experiment we cannot collect sufficient data per subject to assess the consistency of subjects and compute uncertainties from their variability. To do so would require each subject to provide a very large number of responses (in the thousands). Instead we calculate uncertainties for the estimates of texture scale by re-fitting the interval scale to bootstrap re-samplings of our dataset. 1000 re-samplings were used and the estimates from model-fitting to each of these were aligned by their means over the texture set before computing standard deviations across the re-samplings. Figure 3.15 shows the interval scale values and standard deviations; the points are arranged by increasing estimated texture scale. The median standard deviation is 0.29 *jnds*. Four of the textures had much larger uncertainties as there were only a few trials in which a subject judged them as having a smaller scale than another texture. If those trials are missing from a re-sampled dataset then their scale values are poorly constrained by the model.

Figure 3.16 shows a scatter plot of the derived interval scale values for the 112

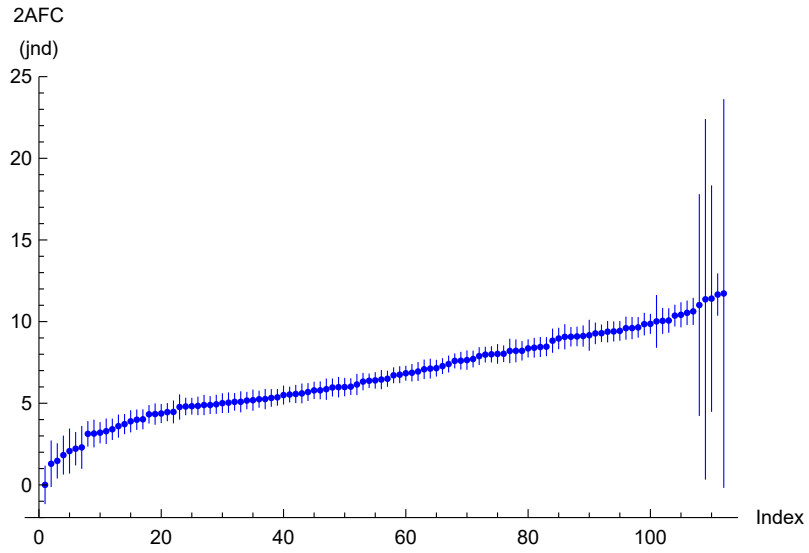


Figure 3.15: 2AFC derived texture scale values (vertical axis) with vertical bars showing a 95% CI of the estimates. The points have been ordered horizontally by increasing texture scale estimate.

Brodatz vs. their standard deviation coloured according to their isotropy score. We noticed that similarly to the first experiment, subjects judged anisotropic textures to have a larger scale than isotropic ones. Another interesting observation was that the standard deviation increased for textures with very small or very large scales compared to those with intermediate scales. This suggests that there was higher agreement between observers on the scale for the latter compared to textures with either very large or very small scales.

3.3.4 Summary

Table 3.4 contains the texture scale estimates derived from the 2AFC study for each of the 112 Brodatz and their corresponding standard deviations.

While we were able to validate our hypothesis on subjects' ability to assess texture scale consistently, irrespective of regularity, in our first perceptual study, the estimates we obtained had wide confidence intervals and the responses were sensitive to subjects' personal systematic bias. In this section we presented a second, 2AFC, study. The responses to this experiment are not sensitive to the aforementioned bias and lead to much tighter confidence intervals. We noticed, similar to the direct-scaling experiment, that subjects judge anisotropic textures to have a larger

Tex	Scale	SD	Tex	Scale	SD	Tex	Scale	SD	Tex	Scale	SD
D1	6.0	0.30	**D29**	3.1	0.42	**D57**	3.2	0.31	**D85**	5.6	0.29
D2	7.1	0.29	**D30**	9.3	0.26	**D58**	11.0	3.45	**D86**	7.6	0.26
D3	4.9	0.28	**D31**	10.0	0.39	**D59**	11.4	5.62	**D87**	6.5	0.28
D4	3.6	0.34	**D32**	2.1	0.69	**D60**	8.2	0.36	**D88**	10.5	0.37
D5	6.5	0.23	**D33**	4.8	0.25	**D61**	9.6	0.30	**D89**	9.4	0.30
D6	1.8	0.59	**D34**	4.4	0.28	**D62**	10.4	0.32	**D90**	10.0	0.81
D7	10.1	0.37	**D35**	5.8	0.25	**D63**	9.1	0.31	**D91**	11.7	0.64
D8	6.7	0.25	**D36**	5.0	0.29	**D64**	6.9	0.28	**D92**	5.4	0.24
D9	4.0	0.30	**D37**	8.0	0.23	**D65**	5.9	0.32	**D93**	5.3	0.30
D10	8.4	0.27	**D38**	4.8	0.38	**D66**	6.0	0.27	**D94**	7.1	0.29
D11	6.9	0.26	**D39**	9.6	0.33	**D67**	6.1	0.31	**D95**	6.3	0.26
D12	7.1	0.24	**D40**	9.8	0.34	**D68**	5.2	0.25	**D96**	9.1	0.29
D13	9.1	0.29	**D41**	9.6	0.34	**D69**	8.8	0.36	**D97**	9.2	0.47
D14	3.1	0.38	**D42**	10.6	0.41	**D70**	8.4	0.29	**D98**	8.0	0.24
D15	7.6	0.29	**D43**	11.7	6.05	**D71**	9.4	0.32	**D99**	9.9	0.29
D16	0.0	0.58	**D44**	11.4	3.52	**D72**	9.3	0.32	**D100**	6.7	0.23
D17	3.9	0.33	**D45**	10.4	0.38	**D73**	6.0	0.27	**D101**	5.2	0.31
D18	7.6	0.22	**D46**	8.0	0.29	**D74**	8.0	0.24	**D102**	5.3	0.26
D19	4.9	0.30	**D47**	8.2	0.34	**D75**	7.3	0.23	**D103**	4.8	0.26
D20	5.5	0.28	**D48**	9.1	0.38	**D76**	5.0	0.30	**D104**	5.1	0.26
D21	1.3	0.71	**D49**	4.5	0.34	**D77**	1.5	0.53	**D105**	5.1	0.31
D22	4.8	0.27	**D50**	7.7	0.24	**D78**	2.3	0.65	**D106**	4.3	0.32
D23	8.4	0.29	**D51**	7.9	0.26	**D79**	3.3	0.38	**D107**	8.5	0.28
D24	4.3	0.27	**D52**	4.0	0.30	**D80**	4.5	0.27	**D108**	9.0	0.32
D25	9.4	0.29	**D53**	2.2	0.50	**D81**	4.9	0.26	**D109**	5.8	0.26
D26	8.2	0.29	**D54**	5.5	0.23	**D82**	5.2	0.26	**D110**	5.6	0.27
D27	7.4	0.25	**D55**	3.4	0.32	**D83**	6.4	0.24	**D111**	6.4	0.21
D28	5.7	0.26	**D56**	6.0	0.24	**D84**	3.7	0.30	**D112**	6.8	0.22

Table 3.4: 2AFC texture scales (in jnds) for Brodatz dataset and their corresponding standard deviations.

scale on average than isotropic ones. We also noticed that there is higher agreement across subjects for intermediate-scale textures compared to large and small-scale textures. The main limitation of the scale values derived through this second experiment is that they are on an interval scale and therefore do not translate directly into physical units.

In the next section (3.4) we compare the two perceptually derived datasets in detail, show that they are in high agreement, and propose a method to translate the texture scale estimates derived from the second study (in jnd units) to pixel units.

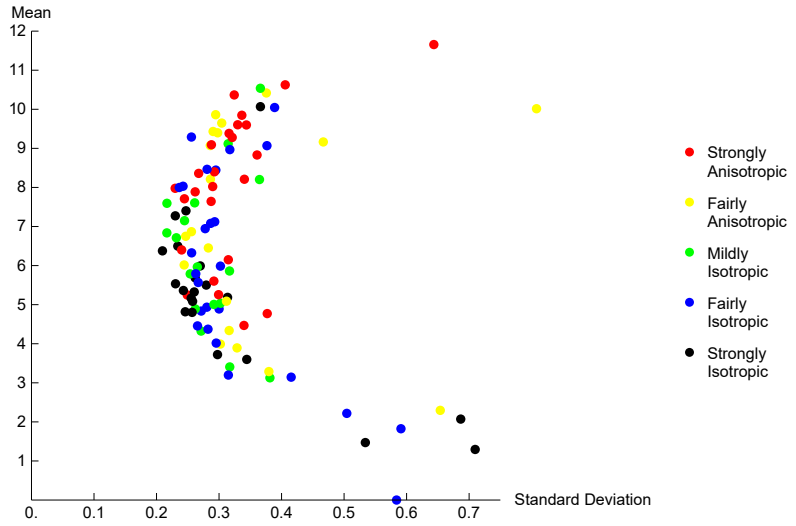


Figure 3.16: Scatter plot of 2AFC-derived scale values *vs.* standard deviation coloured according to isotropy.

3.4 Comparison of the Two Approaches

In order to ascertain that our experiments are measuring a substantive aspect of texture perception, we need to verify that our two experiments are producing consistent estimates. Figure 3.17 shows a scatter plot of the estimates of human perception of texture scale (from the 2AFC experiment) *vs.* the estimates from the window-selection experiment, and their corresponding 95% confidence interval. We see that the relationship between log-window-sizes and the interval scale is strongly linear. The x-axis in 3.17 shows the perceptual texture scale estimates derived from the 2-AFC experiment, which is a discrimination scale. The y-axis shows the texture scale estimates obtained from the direct-scaling experiment. These numbers are actual measurements of scale on the image. Fechner (1966) and Gescheider (1997) show that perceptual estimates derived on discrimination scales obey Weber's Law (which states that increment magnitude detection thresholds are proportional to stimulus magnitude), i.e. $\frac{dI}{I} = c$, where I is the stimulus intensity, and c is a constant. Integrating both sides leads to the log-law for I. Therefore, a linear relationship between the 2AFC interval scale based estimates, and the *log* window sizes from the direct scaling experiment is as expected. The correlation between the two datasets is 0.90, which is a high level of agreement for data of this type. The confidence interval for the correlation can be computed using non-parametric boot-

straps. With 10,000 bootstraps, the resulting 95% CI is $[0.87, 0.94]$. We can also observe, as discussed above, that the estimates arising from the 2AFC experiment have noticeably smaller associated confidence intervals.

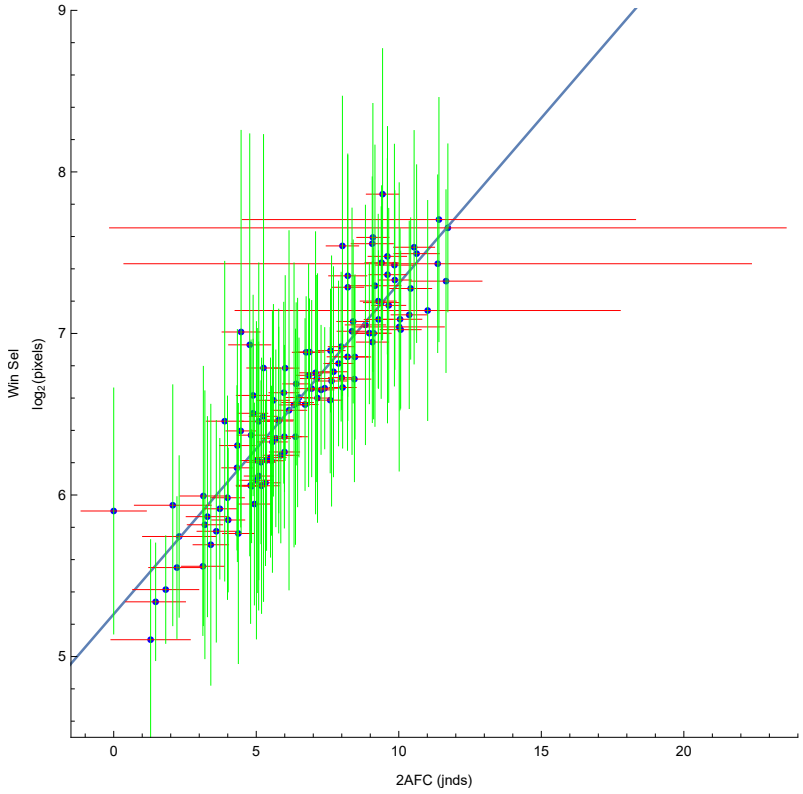


Figure 3.17: Scatter plot of 2AFC-derived interval scale *vs.* log-window size with bars corresponding to 95% CI (in $\log_2$ pixels). Diagonal is $y = x$ lines.

Comparing the two experimental methods we note the following:

1. For window selection, 1792 responses took 3.5 hours of subject time; while for 2AFC, 13,950 responses took 17.5 hours.

2. Subjects reported the 2AFC task to be less taxing.

3. The resulting uncertainties for 2AFC are on average one fifth of those for window selection.

The data from the 2AFC task is useful for obtaining an interval scale with relatively small uncertainties and therefore makes for a good ground truth dataset against which to validate the algorithm for estimating texture scale which we present

in the next section. However, an inconvenience of this dataset is that the units do not represent a physical measure. We therefore used the data from the window selection experiment to derive a linear function that transforms the perceptual interval scale to log-window-sizes, which is easily interpreted. The slope of this mapping allows us to estimate that a change in texture scale (measured in pixels) of 20% is roughly a single *jnd*, hence a doubling in scale is roughly four *jnds*. Table 4.2 shows the scales for the 112 Brodatz textures as derived from the 2AFC task, then linearly transformed to maximally agree with the direct scaling data.

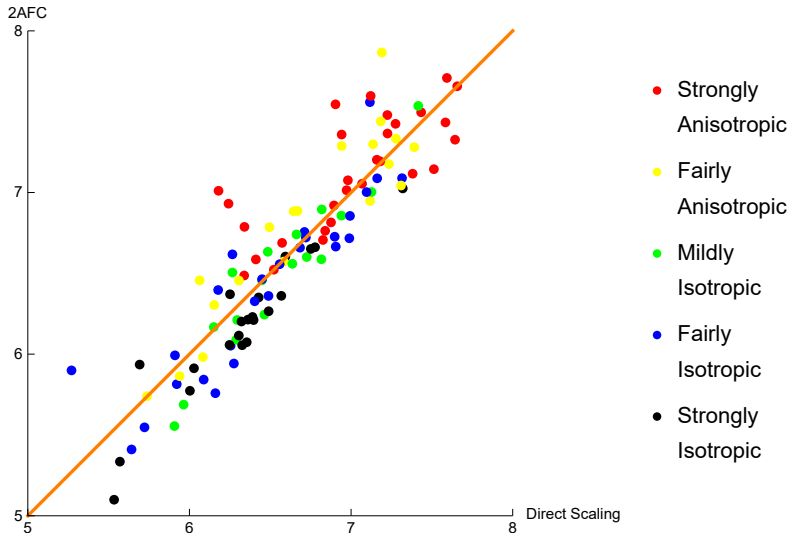


Figure 3.18: Comparison of scales between 2AFC and direct-scaling experiments according to isotropy.

3.5 Summary of Perceptual Studies

In Section 3.2 we presented our first perceptual study. It was a direct scaling study in which subjects were asked to drag and select the smallest upright square window beyond within which texture appears consistently. We restricted the shape of the windows to be upright squares because we wanted *(i)* to force subjects to collapse their assessment to a single value, and *(ii)* to be consistent with the methods of Ardizzone et al. (2013) and Hong et al. (2008). We also observed that there was a substantial amount of per-subject systematic bias leading to wide confidence intervals for the estimated texture scales. Next, we performed a second experiment which did not suffer from this bias, did not impose a restriction on the window

Tex	Scale	SD	Tex	Scale	SD	Tex	Scale	SD	Tex	Scale	SD
D1	6.49	0.06	**D29**	5.91	0.09	**D57**	5.92	0.06	**D85**	6.41	0.06
D2	6.72	0.06	**D30**	7.16	0.05	**D58**	7.52	0.71	**D86**	6.82	0.05
D3	6.27	0.06	**D31**	7.32	0.08	**D59**	7.59	1.15	**D87**	6.58	0.06
D4	6.00	0.07	**D32**	5.69	0.14	**D60**	6.94	0.07	**D88**	7.42	0.07
D5	6.59	0.05	**D33**	6.25	0.05	**D61**	7.24	0.06	**D89**	7.19	0.06
D6	5.64	0.12	**D34**	6.16	0.06	**D62**	7.39	0.07	**D90**	7.31	0.17
D7	7.32	0.07	**D35**	6.45	0.05	**D63**	7.13	0.06	**D91**	7.65	0.13
D8	6.64	0.05	**D36**	6.29	0.06	**D64**	6.68	0.06	**D92**	6.36	0.05
D9	6.08	0.06	**D37**	6.90	0.05	**D65**	6.46	0.06	**D93**	6.34	0.06
D10	6.97	0.05	**D38**	6.24	0.08	**D66**	6.49	0.06	**D94**	6.71	0.06
D11	6.67	0.05	**D39**	7.23	0.07	**D67**	6.52	0.06	**D95**	6.56	0.05
D12	6.73	0.05	**D40**	7.28	0.07	**D68**	6.34	0.05	**D96**	7.12	0.06
D13	7.12	0.06	**D41**	7.23	0.07	**D69**	7.07	0.07	**D97**	7.14	0.10
D14	5.90	0.08	**D42**	7.44	0.08	**D70**	6.98	0.06	**D98**	6.90	0.05
D15	6.83	0.06	**D43**	7.66	1.24	**D71**	7.18	0.06	**D99**	7.28	0.06
D16	5.26	0.12	**D44**	7.60	0.72	**D72**	7.16	0.07	**D100**	6.64	0.05
D17	6.06	0.07	**D45**	7.40	0.08	**D73**	6.48	0.05	**D101**	6.32	0.06
D18	6.82	0.04	**D46**	6.91	0.06	**D74**	6.91	0.05	**D102**	6.35	0.05
D19	6.26	0.06	**D47**	6.94	0.07	**D75**	6.75	0.05	**D103**	6.25	0.05
D20	6.39	0.06	**D48**	7.12	0.08	**D76**	6.29	0.06	**D104**	6.30	0.05
D21	5.53	0.15	**D49**	6.18	0.07	**D77**	5.56	0.11	**D105**	6.30	0.06
D22	6.25	0.06	**D50**	6.84	0.05	**D78**	5.73	0.13	**D106**	6.15	0.06
D23	6.99	0.06	**D51**	6.88	0.05	**D79**	5.94	0.08	**D107**	7.00	0.06
D24	6.15	0.06	**D52**	6.08	0.06	**D80**	6.18	0.05	**D108**	7.10	0.06
D25	7.19	0.06	**D53**	5.72	0.10	**D81**	6.26	0.05	**D109**	6.45	0.05
D26	6.94	0.06	**D54**	6.40	0.05	**D82**	6.32	0.05	**D110**	6.40	0.05
D27	6.78	0.05	**D55**	5.96	0.06	**D83**	6.57	0.05	**D111**	6.57	0.04
D28	6.43	0.05	**D56**	6.49	0.05	**D84**	6.02	0.06	**D112**	6.66	0.04

Table 3.5: Texture scale estimates (in $\log_2$ pixels) for the 112 Brodatz textures, as estimated from the 2AFC experiment, and the corresponding standard deviations of the estimates.

shape, and resulted in much narrower confidence intervals for the texture scale estimates. In Section 3.4, we showed that there is high agreement between the two perceptually derived datasets, as expected from Weber's law.

Given the advantages of the perceptual scales derived from the 2AFC study over those derived from the absolute scaling one, as well as the ease of converting *jnd* units into pixel units for the Brodatz, we choose to use the dataset arising from the 2AFC experiment as ground-truth in the development and validation of the algorithm for estimating texture scale. We present this algorithm in the next chapter.

Chapter 4

Estimating Texture Scale

Our current characterisation of texture scale, "the smallest window size beyond within which texture appears consistently", was based on our observation that non-overlapping domains from a texture look increasingly similar as the size of the domain increases. We illustrated this in Figure 1.2 in the Introduction (Chapter 1). By means of two psychophysical experiments (Chapter 3), we showed that human observers are able to assess the scale of texture consistently irrespective of texture regularity.

In this chapter we present an empirical algorithm for the estimation of texture scale. In fact, we propose two algorithms: one "complete" algorithm that is more accurate than any competing methods in the literature, and a "simplified" approach which is more computationally tractable. We present here the "simplified" approach which is implemented in our online interface to which researchers and the wider public can upload images of textures and obtain estimates of their texture scale. The interface is described in Chapter 5.

We opted for an empirical approach because we wanted to *(i)* remain true to our characterisation and *(ii)* calibrate and validate the model using the texture scale estimates for the Brodatz obtained in our perceptual studies (Chapter 3) as ground-truth.

This chapter is organised as follows. In Section 4.1 we introduce the idea of "texture curves". These texture curves capture the consistency of non-overlapping windows from a texture as the size of the windows varies. To do this we need to

choose *(i)* a representation for texture, and *(ii)* a measure of similarity. We present a few candidate texture representations and justify our decision to use Basic Image Features.

In Section 4.2 we present our work in translating the information captured by our texture curves into texture scale estimates. To do so we need to choose *(i)* what information from the curves we will use (e.g. raw curve values, range, slope, ...), and *(ii)* what approach we will use to convert the selected curve data into scale estimates.

In Section 4.3 we provide a summarised view of our proposed algorithm for estimating texture scale. We also present our more computationally tractable "simplified" approach. For this, we manually select 24 Brodatz textures with clearly discernible texels, segment the texels and record their sizes. Recall that the scale of regular, near regular and homogeneous texture has been characterised by the size of the repeating element from which the texture is composed.

In Section 4.4 we present a comprehensive evaluation of the performance of our algorithm and that of the competition. We look at the consistency of our estimates with the perceptual derived estimates, competing algorithm derived estimates and the sizes of the manually segmented texels.

We conclude the chapter in Section 4.5 with a summary of our proposed method and a discussion of its strengths and limitations.

4.1 Calculating Texture Curves

As mentioned in Chapter 1, we observed that the size of samples taken from a texture appear increasingly similar as the size of the sample increases. We illustrated this in figure 1.2. In this section we present our work on quantifying this observation - a necessary step in the development of an algorithm for estimating texture scale that is true to our characterisation.

Figure 4.1 is a visual representation of our observation regarding the similarity of samples from a texture. As window size increases, the samples will become more similar, or equivalently, less-dissimilar. We call '*texture curve*' a curve measuring

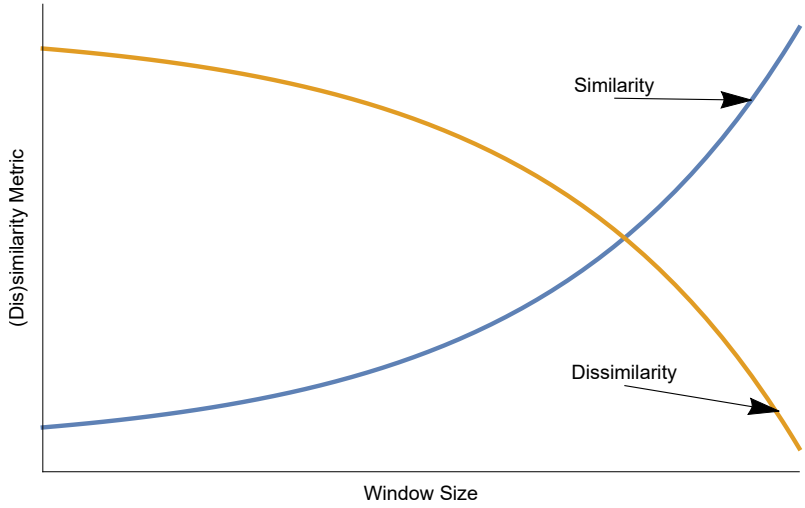


Figure 4.1: Illustration of the concept of texture curves: curves showing variation in similarity or dissimilarity of texture samples with changing window size.

(dis)similarity vs. window size. In this section we present three texture curves that we considered in the development of our texture scale estimation algorithm. In order to calculate these curves we first compute a representation for these textures (Section 4.1.1), and then define a measure of similarity (Section 4.1.2).

4.1.1 Representations

An image is an array of pixels, a grayscale image is an array of illuminance values - we need to choose a way to 'represent' (part of) images. There is a very wide range of image/texture representations available in the literature which are often categorised into a four category taxonomy - statistical, model based, structural, transform based (e.g. Fourier) (Materka et al., 1998).

In this section we present three representations. While we appreciate that there is a much wider choice of texture representations available in the literature and that our choice of three is neither exhaustive nor representative of the wider taxonomy, our choice was carefully made. The first is histograms of gray-scale values. Under this approach, an image region is represented by a histogram of its gray-scale values. We choose to present this approach in Section 4.1.1.1, and trial it in the development of our algorithm because of *(i)* its simplicity, and *(ii)* it is the representation of choice by Hong et al. (2008), one of the two methods against which we compare our algorithm performance in Section 4.4. The second representation (Section 4.1.1.2),

is based on image keypoint density curves. This is the representation adopted by Ardizzone et al. (2013), the second method against which we compare our algorithm performance in Section 4.4. The third representation we present, in much greater detail, in Section 4.1.1.3, is Basic Image Features (BIFs), our image representation of choice. Although there is rich literature on human texture perception, there are no measures that have been validated as capturing human perception for a wide range of textures (Clarke et al., 2011). Therefore, we are restricted to selecting a texture representation that has been proved to be effective at texture retrieval.

4.1.1.1 Histograms of Gray-Scale Values

Recall that in 8-bit images, each pixel can take one of 256 possible values. Each pixel is encoded by a number ranging from 0 to 255 with 0 corresponding to black (not-activated) and 255 to white (fully activated). The Brodatz images are 8-bit grey-scale images. A region of a grey-scale image can be summarised by a histogram of the pixel values. Figure 4.2 shows Brodatz D83 and the grey-scale histogram for a region of the texture.

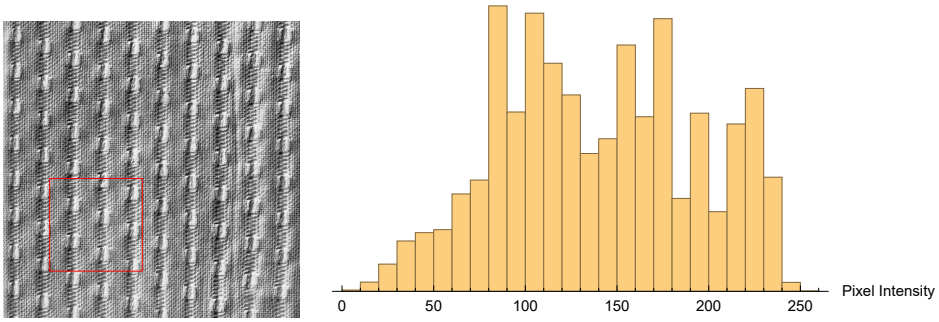


Figure 4.2: Brodatz texture D78 (left) and the histogram of gray-scale values (right) for the highlighted region in the left image.

Grey-scale histograms are easy to compute and summarise an image area into 256 or fewer values which is relatively manageable for image analysis methods. The main limitation, for texture analysis, in using grey-scale histograms to represent image regions is that these histograms do not take into account the structure of the region. All the pixel values are pooled together so the histogram of a circle and a rectangle using the same proportion of grey-levels are exactly the same. Because

structure is crucial for texture, we do not use grey-scale histograms as a texture representation in our work.

4.1.1.2 Keypoint Maps - SURF

Ardizzone et al. (2013) propose a method for identifying texture scale based on the spatial distribution of image keypoints (or points of interest) on the image. At first, this seems like a plausible representation for segmenting tasks, but not necessarily scale detection. However, because we compare the performance of the algorithm we propose in this chapter to that proposed by Ardizzone et al. (2013), we chose to also test a representation similar to theirs.

Image keypoints are defined as 'locations of interest' within an image. They are often used as an intermediary step in local descriptor methods. For each Brodatz we computed the keypoints used in Speeded-Up Robust Features (SURF) (Bay et al., 2006) - the same keypoints used by Ardizzone et al. (2013). Figure 4.3 shows SURF keypoints on Brodatz D83.

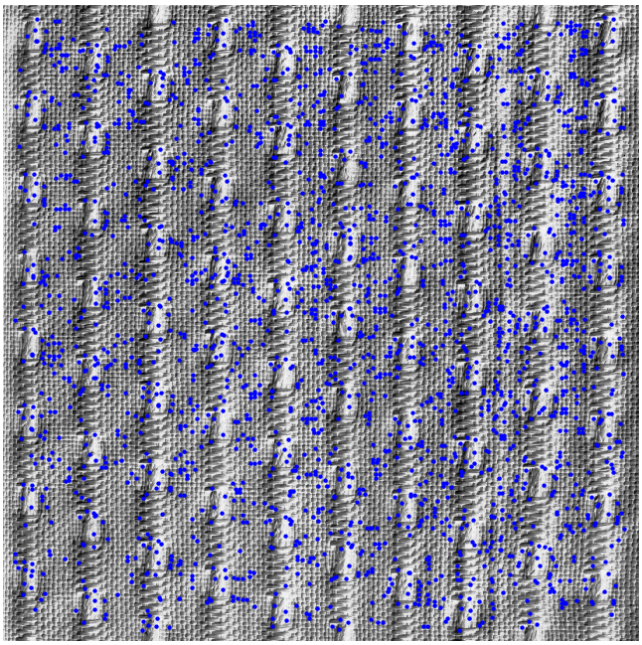


Figure 4.3: SURF keypoints for Brodatz D83 calculated with keypoint strength 0.0006.

The keypoints are found in a three-step process: first by computing the integral image of the texture, second by convolving the integral image with second order DoG kernels by taking its Hessian at each point in the image, and third by comput-

ing the determinant of the Hessian. The determinant of the Hessian at a point in the image represents the 'blobness' of the region around that point. Regions of interest can then be determined by applying a threshold to the resulting matrix of 'blobness'. The two most common types of threshold are the number of keypoints thresholds, and the keypoint strength thresholds. Under the number of keypoints threshold, the pixels with the N highest 'blobness' will be labelled as image keypoints, where N is decided by the researcher. Under the keypoint strength threshold, any pixels with an associated 'blobness' higher than the set threshold will be labelled as a keypoint. Figure 4.3 shows Brodatz D83 with SURF keypoints with strength higher than 0.0006 - the keypoint strength was chosen for visualisation reasons in this case.

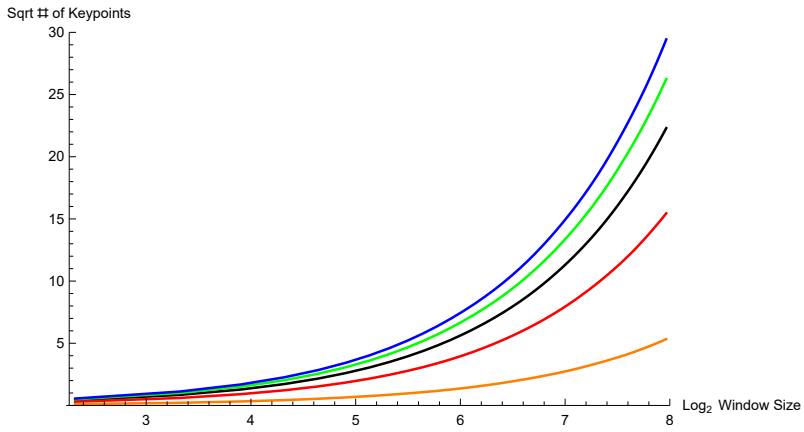


Figure 4.4: Sqrt of number of keypoints (SURF) as we make window size larger for a range of keypoint strengths (0.001, 0.005, 0.0003, 0.0005, 0.0001).

As the window size increases, the number of keypoints in a window will also increase. Figure 4.4 shows the variation in the number of keypoints in upright square windows of varying size for a range of keypoint strengths.

4.1.1.3 Basic Image Features

Basic Image Feature (BIF) histograms (Griffin et al., 2009) are an established and effective representation for recognising texture (Crosier and Griffin, 2010), and even texture segmentation (Jaccard et al., 2017). For a given scale of local image structure, not texture, the BIF representation classifies each pixel into one of seven texton classes according to the symmetry structure of that pixel's neighbourhood. The

classes are flat, slope, light blob, dark blob, light line, dark line and saddle. The classification is based on the output of a family of derivative-of-Gaussian (DoG) filters of the same standard deviation, and orders from 0 to 2. Figure 4.5 shows Brodatz D83 convolved with the derivative of Gaussian filters used in the calculation of BIFs.

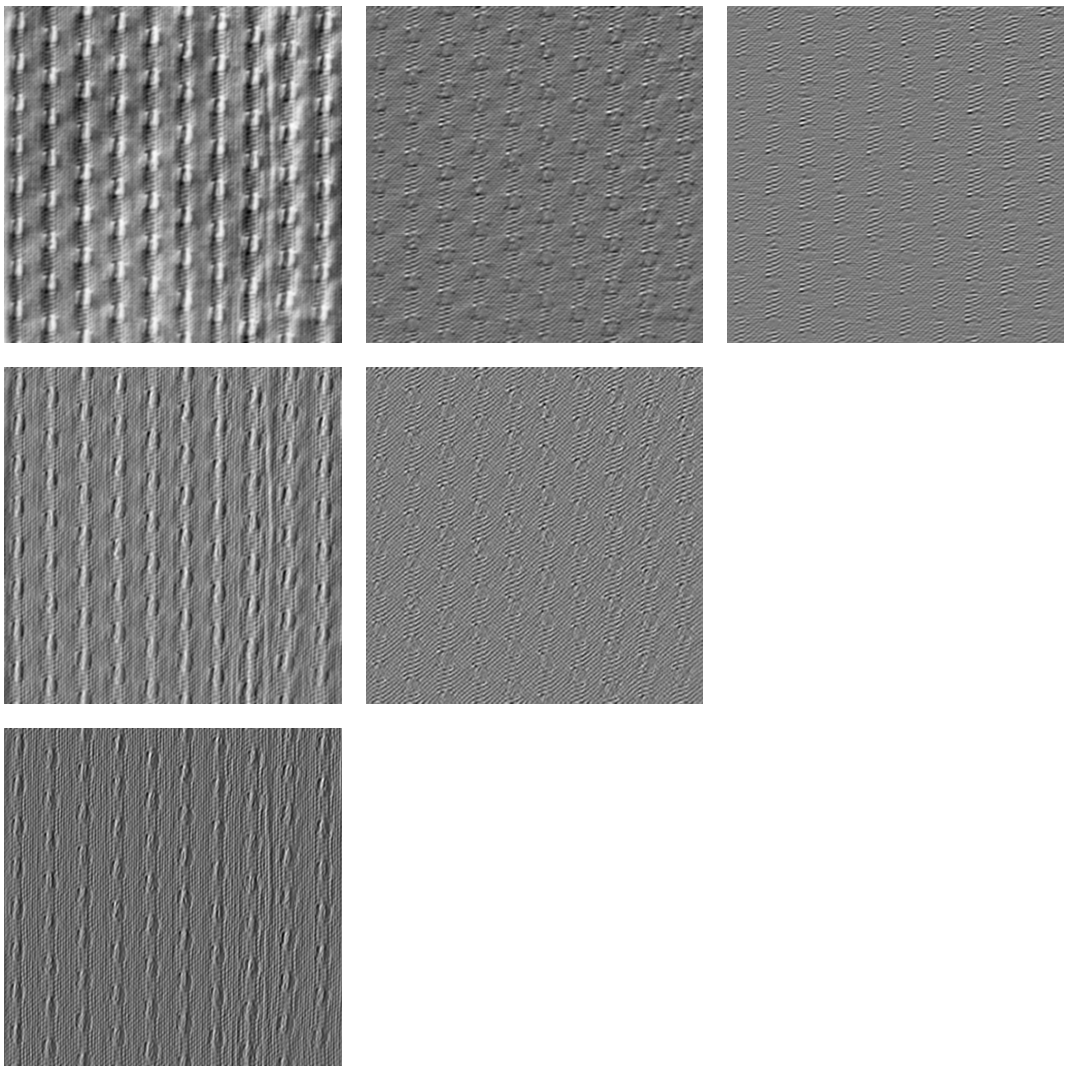


Figure 4.5: Six Gaussian blurs of Brodatz D83 used in the computation of BIF representation. Isotropic blur with $\sigma = 2$ of derivative order 0 to 2 in the x direction (left to right) and 0 to 2 in the y direction (top down).

Let $\{g_{00}, g_{10}, g_{01}, g_{20}, g_{11}, g_{02}\}$ be the set of 2D DoG kernels with variance σ_B^2, of orders zero, one and two. Let $\{c_{ij}\}$ be the convolution of an image I with the DoG kernel g_{ij}, $c_{ij} = g_{ij} \bigotimes I$. The *n-jet* of an image is defined as the set of responses of an image to the convolutions with DoG filters up to order n. Figure 4.5 thus shows

the 2-jet for Brodatz D83.

Griffin (2008, 2009) presents the automorphism groups of scalar functions for both one and two dimensional Euclidean domains. This is the collection of transformations for which a transformed image is indistinguishable from the original structure. For a neighbourhood centred around any pixel, the following numbers may be computed:

- $l = \arctan\left(\frac{c_{20}+c_{02}}{\sqrt{4(c_{10}^2+c_{01}^2)+((c_{20}-c_{02})^2+c4c_{11}^2)}}\right)$
- $b = \arctan\left(\frac{1}{2}\sqrt{\frac{(c_{20}-c_{02})^2+4c_{11}^2}{c_{10}^2+c_{01}^2}}\right)$
- $a = \frac{1}{2}|\arctan\left(2\frac{(c_{01}^2-c_{10}^2)c_{11}+c_{10}c_{01}(c_{20}-c_{02})}{(c_{10}^2-c_{10}^2)(c_{02}-c_{20})+4c_{10}c_{01}c_{11}}\right)|$

The values $\{l,b,a\}$ take values $[-\frac{\pi}{2},\frac{\pi}{2}]\times[0,\frac{\pi}{2}]\times[0,\frac{\pi}{2}]$, and the set of all possible $\{l,b,a\}$ triplets define a bounded three-dimensional manifold. Using an appropriate metric in jet space, this manifold is mapped to the *second order local-structure-solid* (Griffin, 2007), shown in Figure 4.6 (left). Griffin et al. (2009) show that the different highlighted regions on the orbifold correspond to symmetry sensitive areas for seven distinct symmetry types (flat, slope, light blob, dark blob, light line, dark line and saddle).

For single scale BIFs, every pixel in the image is then classified into one of seven texton classes corresponding to the aforementioned symmetry types. The classification algorithm is as follows:

1. Compute the responses to the filter $c_{ij} = g_{ij} \bigotimes I(x)$ for each of the $\{c_{ij}\}$;
2. Compute $\lambda = \sigma^2(c_{20}+c_{02})$ and $\zeta = \sigma^2\sqrt{(c_{20}-c_{02})^2+4c_{11}^2}$;
3. Assign a label between 1 and 7 (corresponding to the seven symmetry types) to each pixel according to which of the following is maximum at that pixel: $\{\gamma c_{00}\ , 2\sigma\sqrt{c_{10}^2+c_{01}^2}\ , \lambda\ , -\lambda\ , \frac{\zeta+\lambda}{\sqrt{2}}\ , \frac{\zeta-\lambda}{\sqrt{2}}\ , \zeta\}$.

The seven classes correspond to regions in the Voronoi cell partitioning (Voronoï, 1908) of the orbifold shown in Figure 4.6 right.

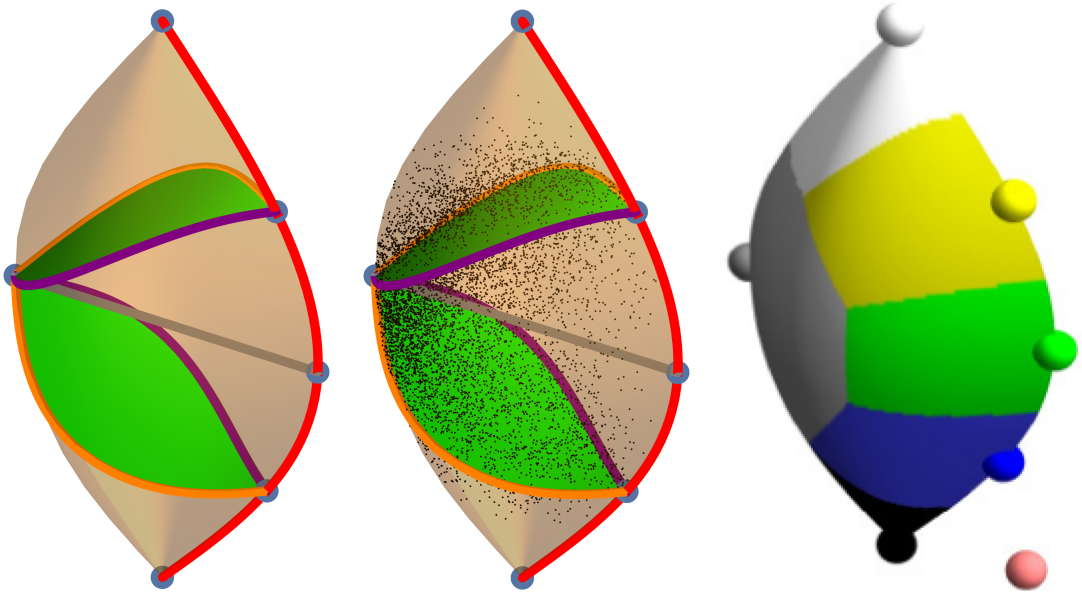


Figure 4.6: Second order solid, with 10,000 points of D83, Voronoi tesselation. Figure from Griffin et al. (2009).

BIFs are controlled by two parameters. The first, σ_{BIF}, is the standard deviation of the Gaussian filters used to probe the local image structure. The second, γ, is a flatness parameter that controls the amplitude of structure that needs to be present, relative to the mean intensity, in order for the local structure to be considered non-flat. We compute the BIF representation for the images at five values of $\sigma_{BIF} = \{0.5, 1, 2, 4, 8\}$ and at five values of $\gamma = \{0, 0.01, 0.02, 0.05, 0.1\}$, hence at 25 BIF parameter settings. It is important to ensure that the BIF representation continues to capture the structure of the texture when choosing BIF parameters. If γ is set too high, high frequency flat regions are not captured, but if it is set too low, too granular regions are labelled flat leading to an increased histogram variance (compare the left hand side to the right hand side columns on Figure 4.7). Too high a σ_B would stop the BIFs capturing structure at the finer level while too low a σ_B would not allow the BIFs to capture structure at the repeating element level (compare top and bottom rows on Figure 4.7). The 25 parameters mentioned above were chosen so as to maximise the amount of structure captured by the BIF representation at all the relevant scales for the Brodatz library. In practice this is done through visual inspection, i.e. we check that the BIF representation of a texture is visually similar

to the actual texture. We have chosen to use Brodatz D83 in Figure 4.7 because it shows the behaviour of BIFs with changing parameters well. Henceforth, we will use the aforementioned 25 BIF paramter settings in all BIF calculations for the Brodatz. We discuss the limitations of making this choice in Section 4.5.

Figure 4.7 shows the BIF representation of Brodatz D83 for 5 different BIF scales (σ_B) and flatness thresholds (γ). Pixels are coloured according to the 7-category BIF labelling: $1 \rightarrow$ Pink (flat), $2 \rightarrow$ Gray (slope), $3 \rightarrow$ Black (dark blob), $4 \rightarrow$ White (light blob), $5 \rightarrow$ Blue (dark line), $6 \rightarrow$ Yellow (light line), $7 \rightarrow$ Green (saddle).

Similarly to grey-scale histograms, an image region is then represented by a 7-bin histogram corresponding to the seven symmetry types, as shown in Figure 4.8. However, BIFs capture the second-order structure of images as each pixel is labelled according to the structure of the area around that pixel, and not simply based on the single pixel value.

In order to capture the second order local structure of an image it is important that the scale of the underlying Gaussian kernels in the BIF classification are appropriate for the image in question. If σ_B is too small, the BIF representation will pick up structure at too small a scale. Similarly if σ_B is too large, the resolution of the BIF representation will be too low to capture the structure at the relevant scale (Figure 4.7). Images may display multiple scales, and often different images within a dateset display structure at different scales. In order to increase the overall classification score in such tasks, the representation needs to capture structure at all these scales. BIFs can deal with this using BIF *columns*. The BIF representation is calculated at c different scales. The scales at which the BIF representations are calculated are given by $\{\sigma_B \times 2^{\frac{i}{2}}\}_{i=1}^{c}$. Every pixel is then assigned c labels between 1 and 7. The pixel labels are combined. Thus every pixel is classified into 1 of 7^c possible texton classes and every image region is represented by a 7^c-bin histogram (Crosier and Griffin, 2010). For example, a 3-column BIF representation classifies the neighbourhood around each pixel according to one of $7^3 = 343$ texton classes. We make use of column BIFs for one of our texture curves in Section 4.1.2. BIF

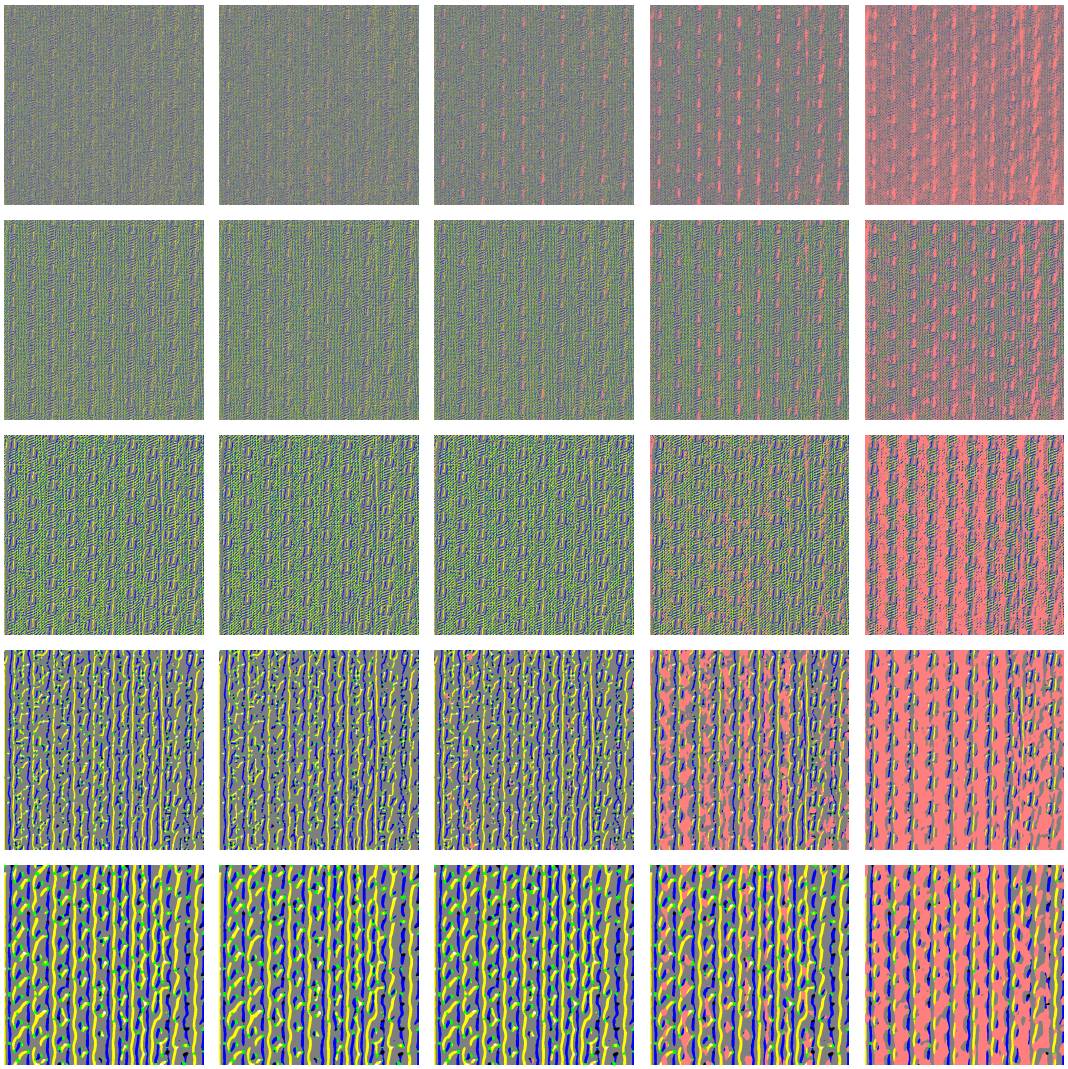


Figure 4.7: BIFs representations for Brodatz D83 for parameter settings σ_B in $\{1,2,4,8,16\}$ top to bottom, and γ in $\{0,0.01,0.02,0.05,0.1\}$ left to right. Colour legend: 1 $\rightarrow$ Pink, 2 $\rightarrow$ Gray, 3 $\rightarrow$ Black, 4 $\rightarrow$ White, 5 $\rightarrow$ Blue, 6 $\rightarrow$ Yellow, 7 $\rightarrow$ Green.

columns can be combined to capture structure at an even larger number of scales within an image. Crosier and Griffin (2010) introduce the concept of *multiscale* BIFs - a collection of BIF columns computed at a range of σ_Bs. The BIF columns for different images are then aligned so as to minimise the distance between the histograms at various scales.

4.1.2 Texture Curves

As mentioned in the introduction to Section 4.1, the purpose of texture curves is to qualitatively capture the increase in similarity of texture regions as the size of

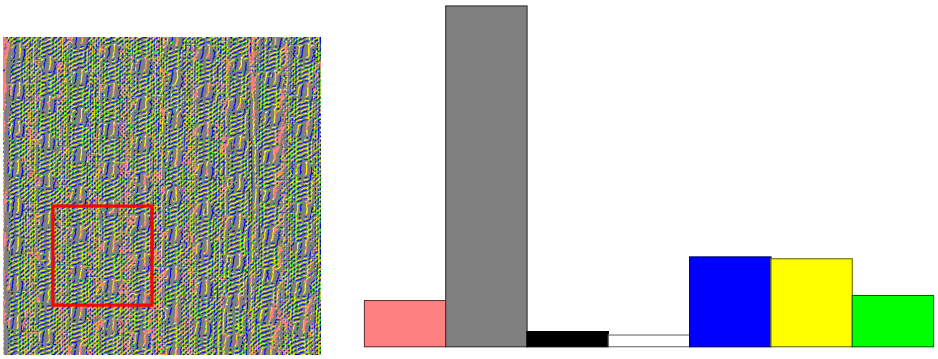


Figure 4.8: BIF histogram (right) for the highlighted region are of the BIF representation of Brodatz D83 (left) calculated using $\sigma_{BIF} = 2$ and $\gamma = 0.05$.

the region increases. There is no established way of capturing this change so we consider three distinct approaches.

First we consider an indirect approach which we call the Odd-One-Out (OOO) approach. In the OOO approach, for a fixed window size, we take two samples from one texture and one sample from a different texture, and record the number of times (as a percentage) that we are able to identify which sample comes from the second texture based on the distances between the BIF histograms of the samples. We present this approach in Section 4.1.2.1.

The second approach we consider is a direct approach which we call the Mean Histogram Distance (MHD) approach. For a fixed window size, we take a large number of non-overlapping sample pairs from a texture, calculate the distance between their BIF histograms, and take the average over a large number of pairs. We present this approach in Section 4.1.2.2.

The first two approaches were selected because they are very true to our characterisation of texture scale - they capture the increase in similarity of BIF histograms for texture sample pairs as the size of the sample increases, as illustrated in Figure 4.9.

The third type of texture curves we present are Total Histogram Variance (THV) curves. They show the total variance of BIF histograms through soft Gaussian windows of varying size. While this approach does not explicitly involve the comparison of sample pairs, it has numerous advantages - it is devoid of the upright

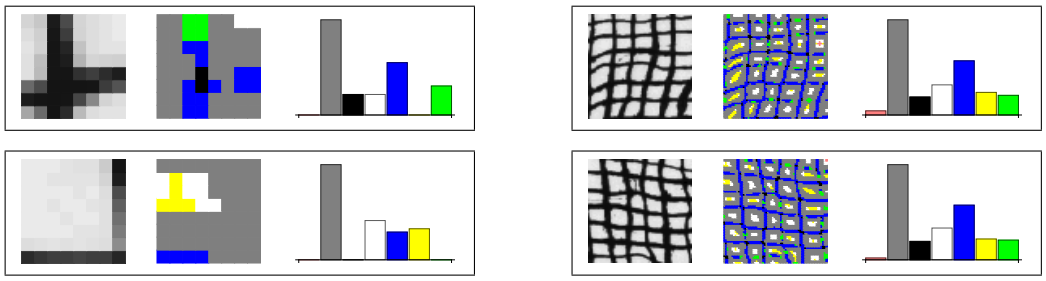


Figure 4.9: Each of the four panels show, left-to-right, a patch from Brodatz D104, its BIF representation, and the BIF histogram. The histograms are shown with square-root transformed bar heights for visibility. The panels on the left are for small patches and the histograms are quite variable. The panels on the right are for larger patches and the histograms are much less variable.

square window restriction of the two previous methods, and is more computationally tractable for larger textures. We present this approach in Section 4.1.2.3.

4.1.2.1 The Odd-One-Out Texture Curves

The OOO algorithm begins as a computational discrimination task. A pair of textures, A and B, are randomly selected from the Brodatz dataset. For a given window-size, two upright square non-overlapping windows, S_1^A and S_2^A, are selected from texture A. One window, of the same size, S_1^B, is selected from texture B. The task consists of identifying which of the three windows, S_1^A, S_2^A or S_1^B, is the odd-one-out (i.e. comes from texture B). This is illustrated visually in Figure 4.10.

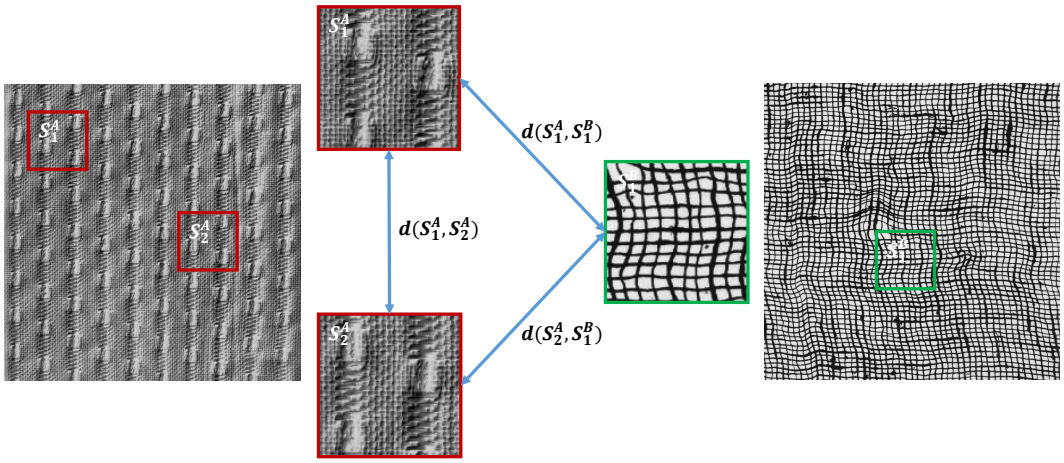
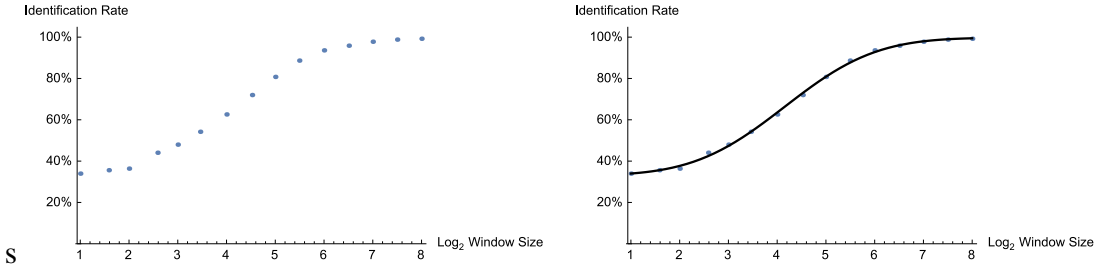


Figure 4.10: Illustration of the OOO method. We have used samples of the images instead of histograms to illustrate that this could have been done for any property of the textures.

Once the windows are selected, BIF histograms are computed for each of the three windows. The pairwise distances between the three sampled windows, $\{d(S_1^A, S_2^A), d(S_1^A, S_1^B), d(S_2^A, S_1^B)\}$, are calculated. If $d(S_1^A, S_2^A) < min\{d(S_1^A, S_1^B), d(S_2^A, S_1^B)\}$ we say that we have correctly identified the odd window. For each texture and each window size, we record the percentage of times we are able to correctly identify the odd-one-out for tasks involving that texture. Figure 4.11 shows the OOO curve for Brodatz D83. We use a Euclidean metric to compute the distances between histograms. We tested a few alternative metrics such as the Bhattacharyya distance (Bhattacharyya, 1943) and a trigonometric distance, and found little to no difference in the shape of the OOO curves resulting from the change in metric.



S

Figure 4.11: OOO curve for the Brodatz D83.

When the window sizes are small, the samples do not contain enough information to allow us to discern the odd texture, there is therefore only a chance probability ($\frac{1}{3}$) of correctly identifying the odd texture. For this reason, many of the curves start around the 33% mark. As the window size increases, windows from the same texture become increasingly similar, as do their BIF histograms. As a result, the proportion of times we are able to identify the odd-one-out increases too and approaches a limit near 100% for sufficiently large windows. Once we have the classification scores for all the window sizes, we fit an error function to the points (see Figure 4.11, right).

Figure 4.12 (left) shows the OOO curves for the 112 Brodatz computed using single scale BIFs and 10,000 iterations. The figure shows that the curves are not monotonic. Increasing the number of trials per texture and window size did not help smooth out the curves. In addition, the calculation for larger windows can

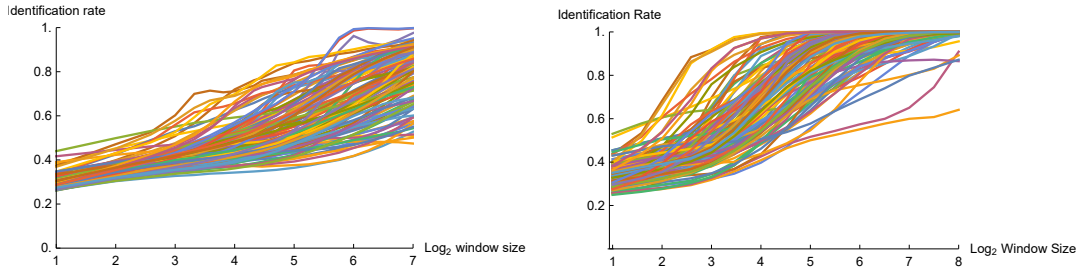


Figure 4.12: OOO curves obtained using single scale (left) *vs.* column BIFs (right). These are the OOO curves for each of the 112 Brodatz averaged over BIF parameter settings.

be computationally expensive so increasing the sample size did not prove to be a viable way of making the curves more monotonic. Instead we opt to use BIF columns over single scale BISs. Specifically we choose BIF columns computed using $\sigma_B = 0.5, 1, 2$ and $\gamma = 0.02$. This leads to monotonic curves for all the 112 Brodatz using the same number of trials (10,000), as seen in Figure 4.12 (right).

4.1.2.2 The Mean Histogram Distance Texture Curves

For each texture in the Brodatz library, and for a given window size, we take $10,000$ pairs of upright square window sample pairs. We then calculate the BIF representation for each of the samples, and calculate the distance between the histogram pairs. The MHD for that texture and window size is the average of the distance between the histograms pairs. Figure 4.13 shows the process visually.

These steps are then repeated for a range of window sizes and for each of the 112 Brodatz. Figure 4.14 shows the MHD vs. $\log_2$ window size for all 112 Brodatz and for 25 BIF parameter settings (σ_B in $\{1,2,4,8,16\}$ top to bottom, and γ in $\{0,0.01,0.02,0.05,0.1\}$ left to right). The shape of the curves is governed by the BIF parameter settings as well as by the texture itself.

Figure 4.15 shows the MHD curves for two Brodatz textures and the corresponding textures. Brodatz D8 (blue) is more regular and the repeating structure happens at a smaller scale than for Brodatz D7 (red). Its corresponding MHD curve sits below that of D7 indicating that, as expected, for a given window size, two samples from D8 are more similar than two samples from D7.

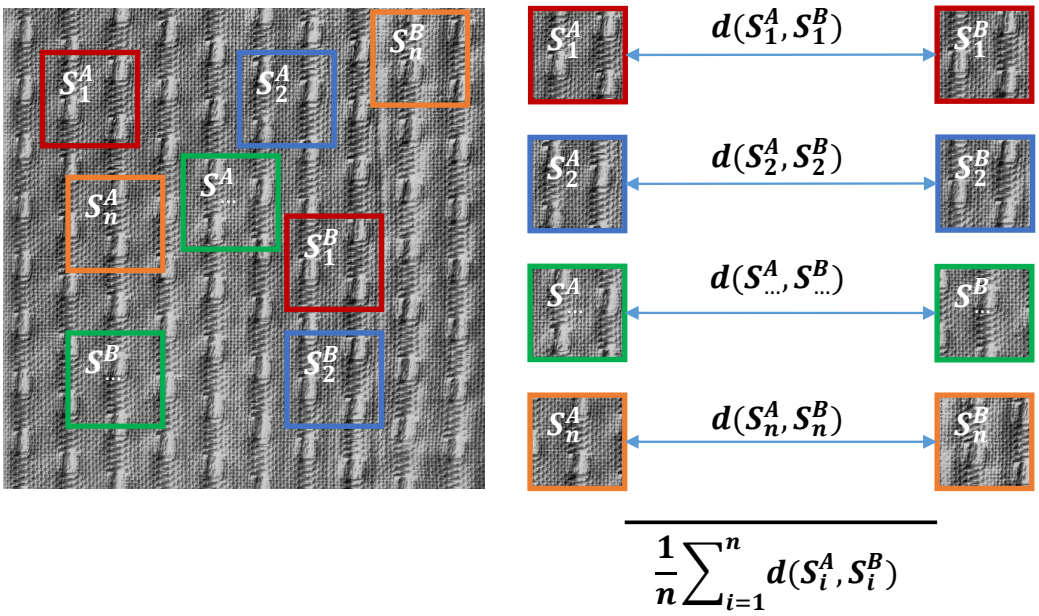


Figure 4.13: Illustration of the OOO method. We have used samples of the images instead of histograms to illustrate that this could have been done for other texture properties.

4.1.2.3 The Total Histogram Variance Texture Curves

The two previously presented texture curves (OOO and MHD) are respectively an indirect, and a direct, set of curves which are true to our characterisation in the sense that they involve comparing window pairs. One problem with comparing window pairs is that the number of samples required to provide good coverage of the texture is very large so the computation of the curves can be slow. In this section we present a set of texture curves obtained from a more 'traditional' measure of consistency/similarity - variance.

Having computed the 25 BIF representations $(5\gamma \times 5\sigma)B$ for each image, we then compute the histograms of all local regions at a range of window sizes. We use soft Gaussian windows whose size we measure by their standard deviation. We use sizes $\left\{2^{\frac{i-1}{4}}\right\}_{i=0...29}$. We call this approach the 'Total Histogram Variance' (THV) approach. Figure 4.16 shows the THV curves for the 112 Brodatz textures for our standard 25 BIF parameter settings.

For any given texture, histograms for small windows will vary more than histograms for larger regions, we therefore expect decreasing curves (as for MHD

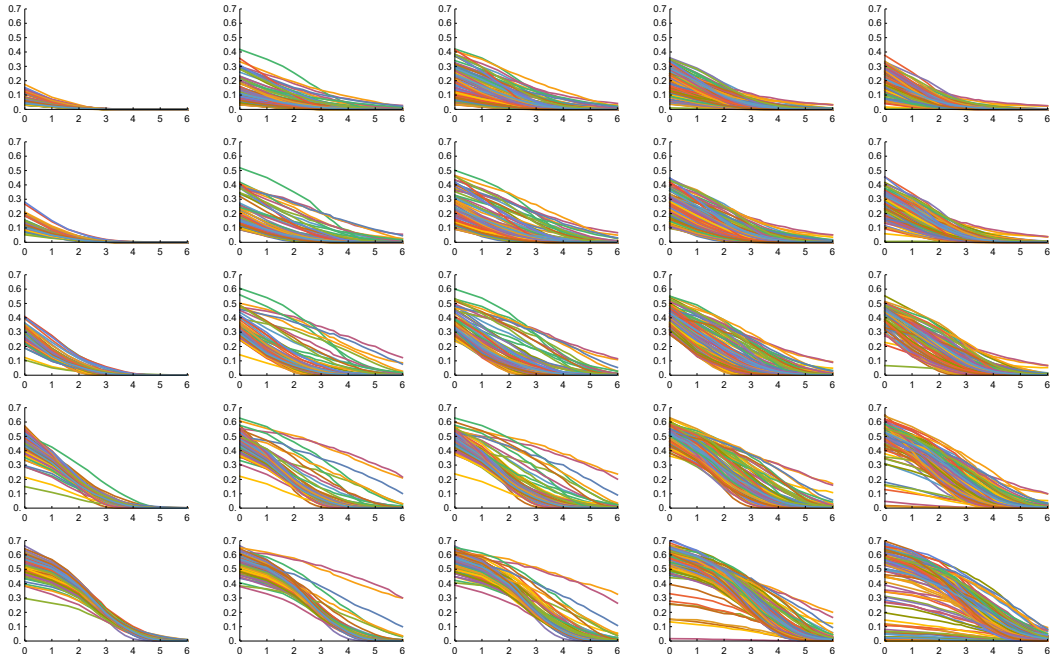


Figure 4.14: MHD curves for the 112 Brodatz for $\sigma_B = 0.5, 1, 2, 4, 8$ (left to right) and $\gamma = 0, 0.01, 0.02, 0.05, 0.1$.

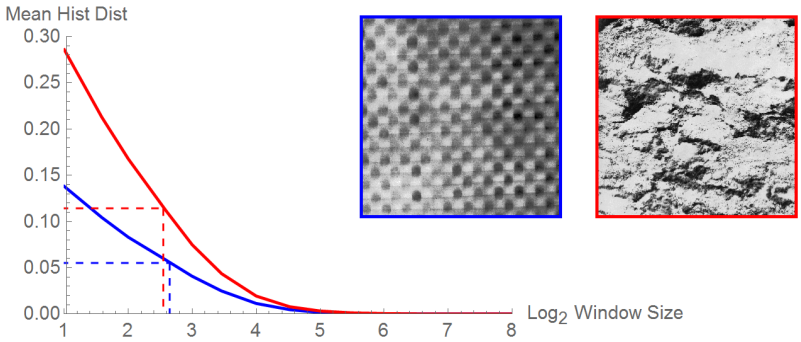


Figure 4.15: Mean histogram distance curves for two Brodatz D7 (red) and D8 (blue). The dotted line illustrates how we obtain a window size estimate when MHD value drops to a certain percentage of that with 2×2 pixel windows, in this case 60%.

curves). Figure 4.17 shows the 25 TV curves, one for each of our BIF settings, for two Brodatz textures. We notice that there is little variation in the curves for the more regular and uniform texture (D102, Figure 4.17 bottom) than for the less regular and uniform texture (D38, Figure 4.17 top). One may observe that the curves appear smoother than the OOO and MHD curves. This is due to two reasons. First, by design, THV curves are computed for all regions in the texture, while the other

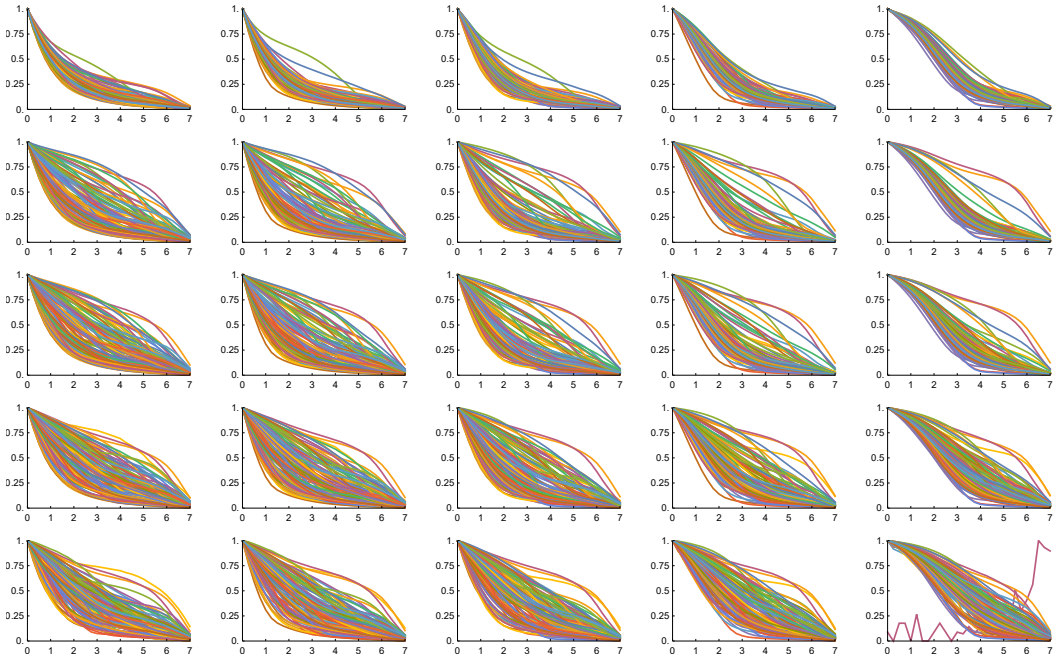


Figure 4.16: These are the total variance curves for the 112 Brodatz across 25 BIF parameter settings ($5\sigma_{BIF} \times 5\gamma$). The have been rescaled so that the variance for single pixels is 1.

two methods are computed for a uniform sample of windows on the texture. Second, because of the increased computational tractability, we sample window sizes closer to one another (in multiplicative steps of $2^{0.25}$ pixels) compared to the other two methods (in steps of $2^{0.5}$ pixels).

To capture the variability of the local histograms we sum the variances (across histograms from windows at different positions) of each of the seven bins. This is equivalent to computing the 7×7 covariance matrix of the histograms and calculating its trace, which is equal to the sum of its eigenvalues, but is simpler and quicker. The sum of the eigenvalues of a covariance matrix is a measure of total histogram variance. For each texture, and every window size, we compute the total variance of the local histograms and plot them against window size. The curves are linearly rescaled in order to make the variance of the histograms for windows of 1×1 pixel equal to 1. The reasoning behind this is that subjects would find it equally difficult to tell whether two individual pixels came from a given texture, irrespective of the texture.

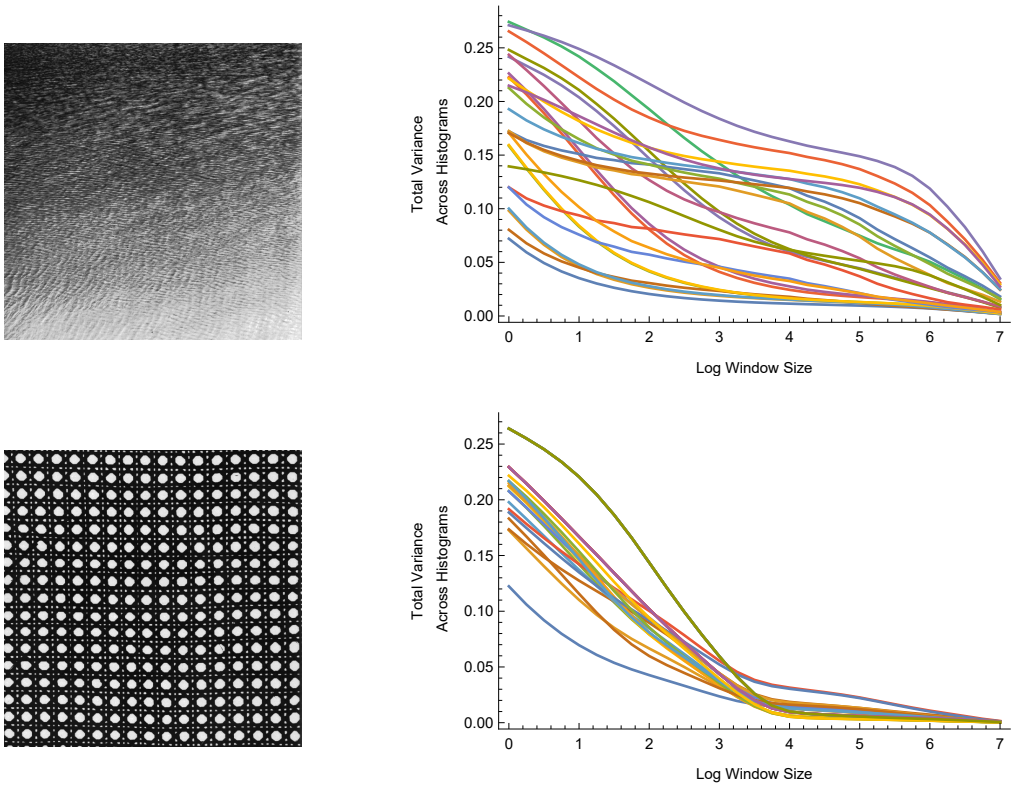


Figure 4.17: Two Brodatz textures (D38, top, and D102, bottom) and their corresponding THV curves for the 25 BIF settings.

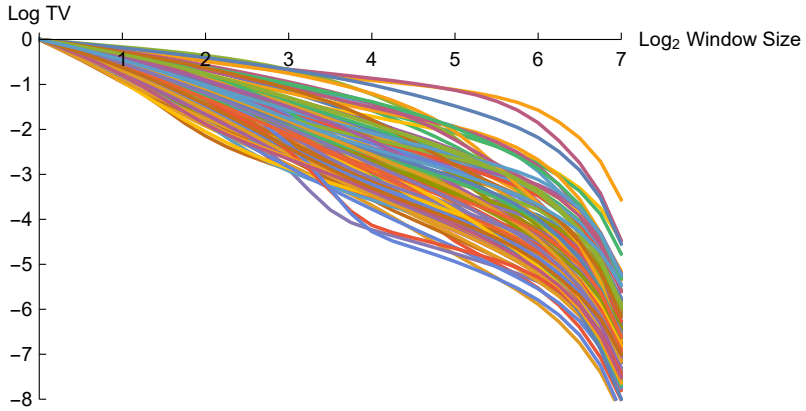


Figure 4.18: Log of the mean (across parameter settings) rescaled THV curves for the 112 Brodatz textures.

While the texture curves for the different Brodatz look different for different textures, the variance for large windows is similar across textures, as can be seen in Figures 4.16 and 4.17. When attempting to estimate texture scale using texture curves, we found that the accuracy of our estimates was poorer for textures with large repeating patterns. We attributed this to the compactness of the texture curves

on the large window size end of the curves. In order to correct for this, in practice we use log-THV curves, as shown in Figure 4.18. Taking the logarithm of the curves widens the large values for larger window sizes thereby improving the texture scale estimates for textures with a larger repeating pattern.

4.1.2.4 SURF Keypoint Curves

As an alternative to BIF-based curves, we used curves showing how the average number of SURF keypoints values with window size for a range of keypoint strengths. For a given Brodatz, the curves are calculated as follows:

- Calculate the locations of SURF keypoints;
- Count the number of keypoints found inside a window of size w as it slides across the image;
- Repeat for window sizes $\left\{2^{\frac{i-1}{4}}\right\}_{i=0...29}$.

Figure 4.19 shows the keypoint curves for each of the 112 Brodatz at the 4 parameter settings $0.0001, 0.001, 0.002, 0.005$ for window sizes $\left\{2^{\frac{i-1}{4}}\right\}_{i=0...29}$. When the keypoint strength threshold is set to the lowest value we consider, 0.0001, the curves are relatively similar to each other (top left of Figure 4.19). As we increase the keupoint strength threshold to 0.005, the curves for the Brodatz become more spread out (bottom right of Figure 4.19).

4.1.2.5 Discussion on Texture Curves

In Section 4.1.2 we presented three types of texture curves, OOO, MHD and THV, which capture the variation in similarity/consistency across window size. The OOO and MHD curves are more 'true' to the observation that motivated our characterisation - samples from a texture look increasingly similar as the size of the window increases. The OOO curves are an indirect measure of similarity as they do not show distance or variance, instead they show the percentage of times for which the content of the samples enables us to identify the odd one out. The MHD method is a direct approach showing how the mean distance between BIF histograms of samples from a texture vary according to the size of the sample. While these curves

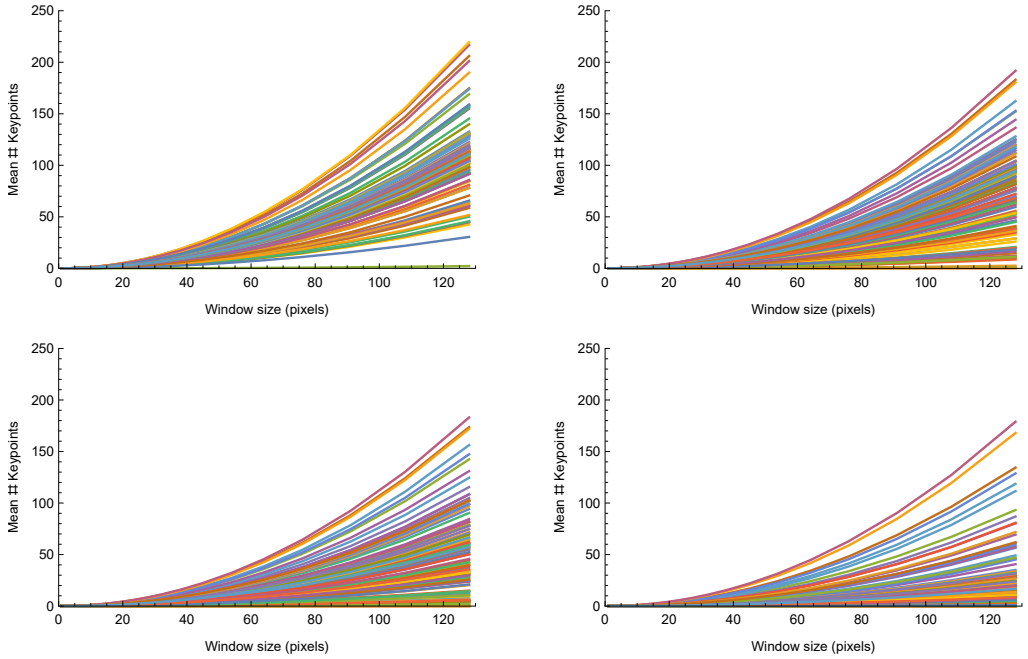


Figure 4.19: SURF keypoint density curves calculated for parameter settings $0.0001, 0.001, 0.002, 0.005$ left to right, top to bottom.

are more aligned to our initial hypothesis, they are more computationally expensive, especially for larger window sizes, than the THV curves.

The third type of curve we presented are the THV curves. These curves show the variance across all histograms across the texture for soft Gaussian windows of different sizes. This method has multiple advantages over the OOO and MHD methods. Firstly, it uses soft Gaussian windows instead of upright square windows, which is more alike to a receptive field in human vision. Secondly, all windows within a texture are taken into account, therefore THV curves capture variance across the entire texture, as opposed to a subset of windows across the texture like the two other methods do. Thirdly, variance is a widely accepted measure of consistency. Finally, it is less computationally expensive thus allowing us to obtain smoother curves through more granular window size sampling. The additional computational tractability allows us to make this algorithm available through an online calculator which we present in Chapter 5.

The THV curves will be our texture curve of choice going forward. We compared the accuracy of the estimates that can be obtained using the three types of

curves and found that there is a small performance advantage in using THV curves over OOO or MHD curves.

4.2 Algorithm for Estimating Texture Scale

Having converted our qualitative description of increasing similarity in texture scale to quantitative texture curves, we have reduced the problem of obtaining texture scale estimates from textures to one of obtaining texture scale estimates from texture curves. In order to do so, three decisions need to be made:

1. **Choice of curve summary:** What information from the texture curves should be passed to the empirical algorithm? We discuss the various options and our choice in Section 4.2.1

2. **Choice of mapping method:** How do we map the curve summary values into the perceptual ground truth? We discuss some candidate maps and assess their merits and limitations in Section 4.2.2

3. **Performance measure:** How do we assess the performance of our method and competing methods? We discuss a few candidate methods, and the care that needs to be taken when evaluating the performance of a method on the same dataset on which it is trained in Section 4.2.3

In Section 4.3 we will provide a summarised view of the algorithm for estimating texture scale, as well as a simplified version of the same algorithm implemented in the online texture scale calculation (presented in Chapter 5) which is more computationally tractable than the complete algorithm.

4.2.1 Choice of Curve and Data

In Section 4.1 we presented three types of texture curves: OOO, MHD and THV. There were two primary objectives in computing the texture curves. The first was to obtain a quantitative method to assess our observation that samples from a texture become increasingly consistent as the size of the samples increases. The second was to reduce the task of deriving texture scale estimates from the image to the

task of deriving texture scale estimates from the texture curves which encode our characterisation.

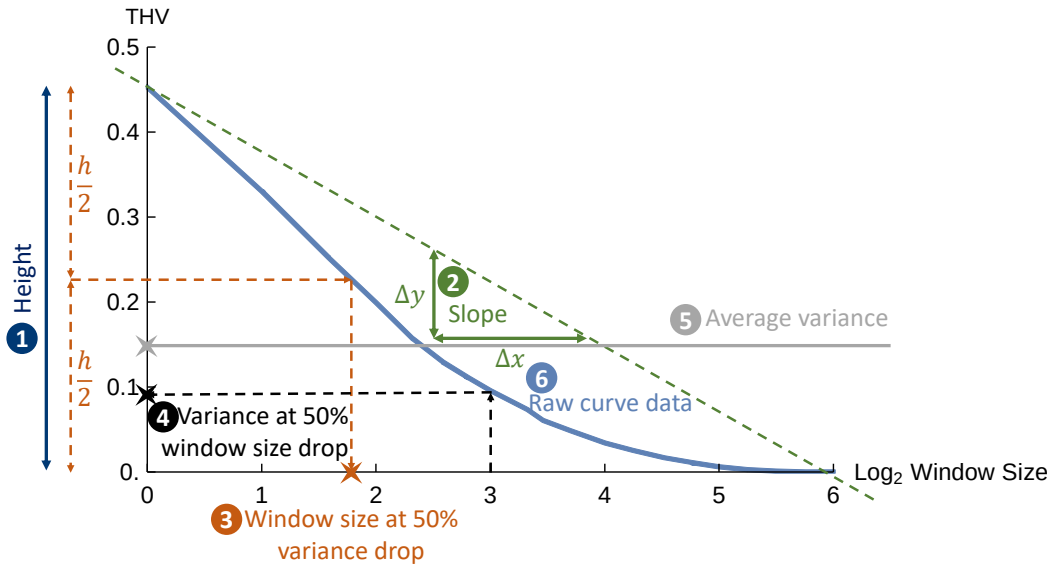


Figure 4.20: Candidate metrics from texture curves.

Figure 4.20 shows six candidate metrics on a sample THV texture curve that can be used to map to the perceptual ground-truth. They are:

1. **Height:** the height of the curve, i.e. the mean variance for histograms for single pixel regions

2. **Slope:** the slope of the texture curve taken between the 1×1 pixel windows and the window size at which the total histogram variance drops by 95% of what it was for single pixel regions

3. **Window size at 50% variance drop:** the size of the window at which the variance drops to 50% of what it was at the smallest window size

4. **Variance at 50% window size drop:** the variance of histograms for window sizes half-way between the smallest window size and the window size at which the variance drops by 95% of what it was for single pixel regions

5. **Average variance:** the average variance across uniformly sampled window sizes

6. **Raw curve data:** the actual values on the curves can be passed to a regression function, or similar, without any further pre-processing

We tested all of these options and a few more depending on the map we used. Ultimately, for the sake of simplicity, we chose to use a linear regression map and the raw curve data. A further consideration that needs to be made is regarding how to deal with the 25 curves obtained for each texture, one for each of the BIF parameter settings. For our complete method we chose to use all the points from all 25 curves and let the regression function decide what to do. For the simplified methods, we average the curves across BIF settings and select a single point from the curve, reducing the computational complexity.

4.2.2 Translating Curve Data to Texture Scale Estimates

Once we have made a choice as to what data from the curve we use to map to the perceptual ground-truth values, we need to make a decision as to which function to use for the map. Figure 4.21 shows a scatter plot of the midpoints of the average (across BIF settings) log THV curves. The relationship between the raw curve data is approximately linear. We have overlaid four types of functions we have used to map the THV curve data to the perceptual ground-truth:

1. **Linear regression:** linear regression line fitted through the scatter plot standard least-squares method.

2. **Quadratic regression:** a quadratic function is fitted through the scatter plot using the standard least-squares method.

3. **k Nearest Neighbours regression:** the value at a point was estimated by the the Euclidean average of the 10 nearest neighbours. Lowering the k would result in a curve with more pronounced spikes.

4. **Random Forest: regression:** the value at a point is defined by an n part piecewise linear function. The example shown in Figure 4.21 was obtained using a 40-branch decision tree. We considered setting the number

of branches manually but found that Mathematica worked well for selecting a suitable number of branches.

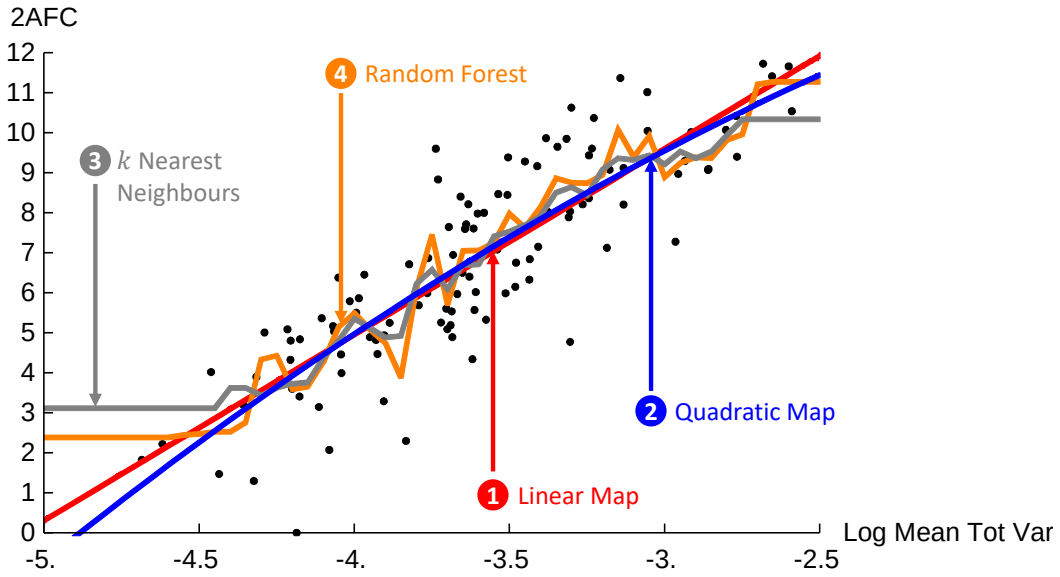


Figure 4.21: Candidate maps from curve data to perceptual ground-truth.

The points illustrated in Figure 4.21 are log-transformed values. This transformation not only allows us to spread the variances for large window sizes (as discussed in Section 4.1.2.3) but also makes the relationship between the midpoint of the curves and the perceptual ground-truth linear. For the OOO curves introduced in Section 4.1.2.1 we also used an error function (erf) mapping to the OOO curves.

In addition to these approaches, we considered using supervised machine learning approaches such as deep learning but decided not to do so for the following two reasons. First, our dataset is small compared to the datasets required for deep learning methods to be successful. Second, the methods we are developing are intended to be used by life scientists who wish to understand how they work, and, more importantly, when they do not work which is not always understood in deep neural networks (Tanay and Griffin, 2016).

4.2.3 Performance Assessment

We assess the performance of the estimates obtained using both our, and competing, methods. To do so, we report RMSE between the algorithm-derived estimates

and the perceptual ground-truth from the 2AFC dataset to the Brodatz collection. Recall that the texture scale ground-truth measurements obtained for the Brodatz through the 2AFC experiment are on an interval scale spanning 12jnds. A span of 12jnds indicates that the visual system is able to discern 12 levels of scale within the Brodatz library. The number of scale levels that an algorithm can distinguish can be estimated by dividing the span of the interval scale by the RMSE of the estimates. Thus, an RMSE of 1jnd would indicate that an algorithm has a similar sensitivity to texture scale which is similar to the visual system. This is the benchmark for which algorithms designed to be practical aim.

Given that the parameters for the maps between the texture curve data and the perceptual ground-truth values are tuned on the Brodatz, and that the performance assessment is also carried out on the Brodatz, we perform a Leave-One-Out (LOO) cross-validation approach to ensure that the RMSEs are not biased. To do so, we exclude one texture from the Brodatz, and 'train' the map on the remaining 111 Brodatz textures. We then obtain a texture scale estimate for the excluded texture. We repeat this process in turn for each of the 112 Brodatz. The result of doing this is that none of the texture scale estimates are obtained for images already included in the training set.

In addition to reporting the RMSE, we report the correlation between the perceptual ground-truth datasets, the estimates obtained using our algorithm, those obtained using competing algorithms, and the texel size for regular textures.

4.3 Algorithm for Estimating Texture Scale

Having discussed multiple choices of image representation, types of texture curves, curve metrics, and mapping functions, we converged on a preferred method that offers a good balance between performance and simplicity. In this section we present a summarised view of our proposed algorithm for the estimation of texture scale. In Section 4.3.1, we provide a step by step guide to our complete method (using BIFs at multiple BIF settings), which is computationally tractable on a standalone computer and offers competition-beating texture scale estimates performance. In

Section 4.3.2, we present a simplified method which is around $10\times$ less computationally expensive but $70-80\%$ as accurate.

4.3.1 Complete Method

We propose the following algorithm for obtaining texture scale estimates that are consistent with human assessment:

1. **Pre-process:**

   (a) Convert image to gray-scale: if not already gray-scale

   (b) Resize: the BIF parameters used in our algorithm are optimised for images of a size of around 640×640 pixels. If the image you wish to process is much larger or smaller than this, the image needs to be resized or cropped. When resizing/cropping, the repeating elements (if present) need to remain discernible (i.e. not too small) but not be larger than $\frac{1}{4}$ of the image

2. **Compute THV curves:**

   (a) Compute the BIFs at 25 parameter settings:(σ_B in $\{1,2,4,8,16\}$ top to bottom, and γ in $\{0,0.01,0.02,0.05,0.1\}$)

      i. Compute the responses to the filter $c_{ij} = g_{ij} \bigotimes I(x)$ for each of the $\{c_{ij}\}$, where g_{ij} is the a DoG kernel with mean 0 and standard deviation σ_{BIF} of order i in the $x-$direction and j in the $y-$direction;

      ii. Compute $\lambda = \sigma^2(c_{20}+c_{02})$ and $\zeta = \sigma^2\sqrt{(c_{20}-c_{02})^2+4c_{11}^2}$;

      iii. Assign a label between 1 and 7 (corresponding to the seven symmetry types) to each pixel according to which of the following is maximum at that pixel:
      $\{\gamma c_{00}\ , 2\sigma\sqrt{c_{10}^2+c_{01}^2}\ , \lambda\ , -\lambda\ , \frac{\zeta+\lambda}{\sqrt{2}}\ , \frac{\zeta-\lambda}{\sqrt{2}}\ , \zeta\}$.

   (b) Calculate the total variance curves at 29 window sizes, $\{w_i\} = \left\{2^{\frac{i-1}{4}}\right\}_{i=0\ldots29}$ pixels, for each of the 25 BIF representations:

i. Convert the BIF representation, $\mathbb{B} - ij$ (a matrix with labels 1-7 for each pixel representing texton classes corresponding to the predominant symmetry type in the neighbourhood surrounding said pixel) into 7 binary layers, $\{B^k_{i,j}\}_{k=1}^{7}$, one for each texton class;

ii. Blur the binary layers: convolve each of the binary layers with a Gaussian kernel of variance $\frac{1}{2}w_i$;

iii. The total variance is then calculated as the average of the variances of each blurred binary BIF layer. This is equivalent to computing the 7×7 covariance matrix of the histograms and calculating its trace, which is equal to the sum of its eigenvalues. The sum of the eigenvalues of a covariance matrix is a measure of total variance.

(c) Repeat: repeat for each of the 25 BIF settings

(d) Rescale and take logs: rescale the raw THV curves so that the variance for the smallest window size is equal to 1, and take logs of the values

3. **Linear transform:** pass the 725 curve values (25 BIF settings $\times$ 29 window sizes) through a linear regression

4. **Convert to pixels (optional):** the estimates produced by our algorithm are in jnd units but these can be converted into pixels as described in Section 3.4

Estimates using this approach take approximately 3 minutes to compute for a 640×640 pixel gray-scale texture using Mathematica on a mid-price Quad-Core laptop with a processor running at 3GHz.

4.3.2 Simplified Method

The simplified method has two small differences compared to the complete method:

- **Single window size:** we only compute the BIFs at one window size ($\sigma_B = 4$) instead of at the 29 window sizes used in the complete method

- **Single descriptor regression:** instead of passing the THV values at the 25 different window sizes to the regression function, we take the average over

the 25 BIF settings and pass a single point per texture to the linear regression function

In Section 4.4 we will find that the simplified method is $70-80\%$ as accurate as the complete method. Despite this, it is approximately $10\times$ less computationally expensive - estimates using this approach take approximately 20 seconds to compute for a 640×640 pixel gray-scale texture using Mathematica on a mid-price Quad-Core laptop with a processor running at 3GHz.

4.4 Results and Performance

We assess the performance of our algorithm by reporting the root mean squared error (RMSE) between the algorithm predictions and the 2AFC human perception dataset. All reported RMSE values were calculated using a leave-one-out approach for the regression mapping the texture curves to the perception-derived texture scale estimates. We also report 95% confidence intervals for these values, which were obtained by bootstrapping over the images in the sets. For each bootstrap, we repeat the leave-one-out estimates and calculate a new RMSE. These confidence intervals thus indicate the precision of our performance estimates given the finite collection of textures used for evaluation. An RMSE of around 1.0 jnds will indicate that our model predictions agree with the psychophysical data roughly at the precision level of human judgement. Figure 4.22 shows a scatter plot of the estimates obtained using both our complete algorithm (left) and simplified algorithm (right) *vs.* the perceptual ground-truth from the 2AFC experiment. We see good overall agreement between the estimates obtained using both methods. The points for the complete method are closer to the $y = x$ line indicating, as expected, better quality estimates.

One of the main differentiating factors of our algorithm (aside from perceptual validation) is that it is designed to work for textures of all regularities. In order to check whether our algorithm is behaving correctly for all types of textures, we have manually labelled the Brodatz according to one of three regularity classes: 'Fairly regular', 'Highly irregular', or 'Neither'. Figure 4.23 shows the scatter plot of our complete algorithm estimates *vs.* 2AFC ground-truth where the points, correspond-

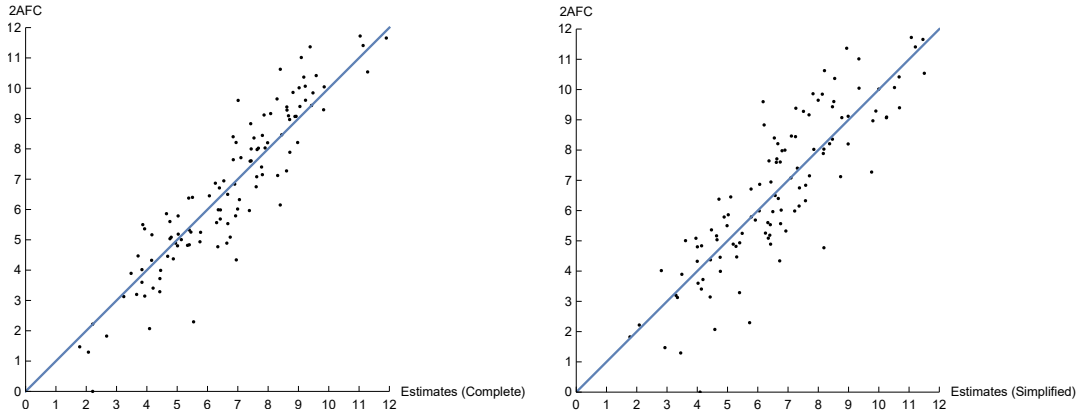


Figure 4.22: Scatter plot of algorithm derived scale *vs.* 2AFC ground-truth. The left scatter plot shows the estimates derived using the complete version of our algorithm, and the right scatter plot shows the estimates derived using the more computationally tractable simplified version of our algorithm.

ing to each of the Brodatz, have been coloured according to this three-category labelling. We are happy to see that there is no clear pattern indicating that our predictions are any better or worse for regular textures compared to irregular textures, and thus are satisfied that our algorithm performs well irrespective of texture regularity. We do notice a minor trend in that there are more blue points, corresponding to irregular textures, above the $y = x$ lines than below it, indicating that our algorithm has a mild tendency to underestimate the scale for such textures compared to human assessment.

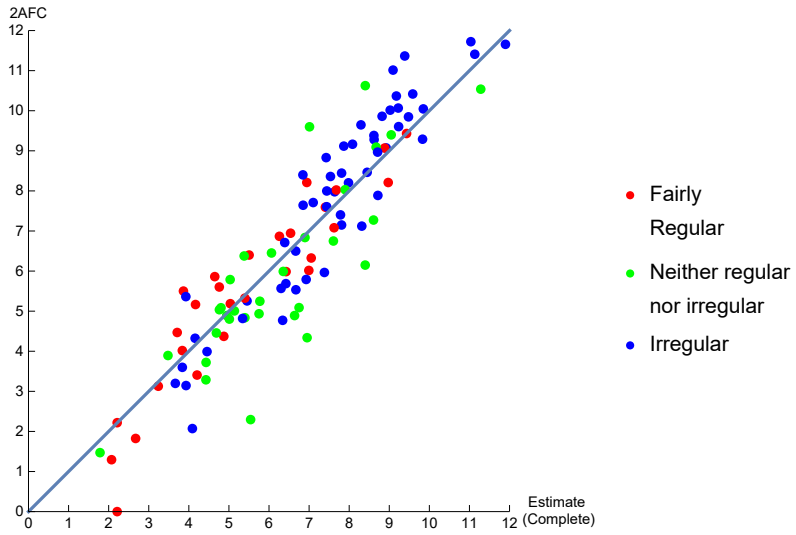


Figure 4.23: Algorithm estimates *vs.* 2AFC coloured according to regularity.

Another potential concern with our approach is that we use Gaussian or square windows which are isotropic, but the repeating elements forming a texture are not always isotropic. To check for any issues related to isotropy, we manually categorise each of the Brodatz into one of five isotropy classes: 'Strongly Anisotropic', 'Fairly Anisotropic', 'Mildly Isotropic', 'Fairly Isotropic' and 'Strongly Isotropic'. Figure 4.23 shows the scatter plot of our complete algorithm estimates vs. 2AFC ground-truth where the points, corresponding to each of the Brodatz, have been coloured according to this five-category labelling. While we are pleased that we do not observe a pattern of one type of texture being further away from the $y = x$ line than another, there are more red and yellow points (anisotropic) above the $y = x$ line and more blue and black (isotopic) points below the $y = x$ line. This indicates that our algorithm has a tendency to underestimate the scale for anisotropic textures and overestimate the scale of isotropic ones compared to human assessment.

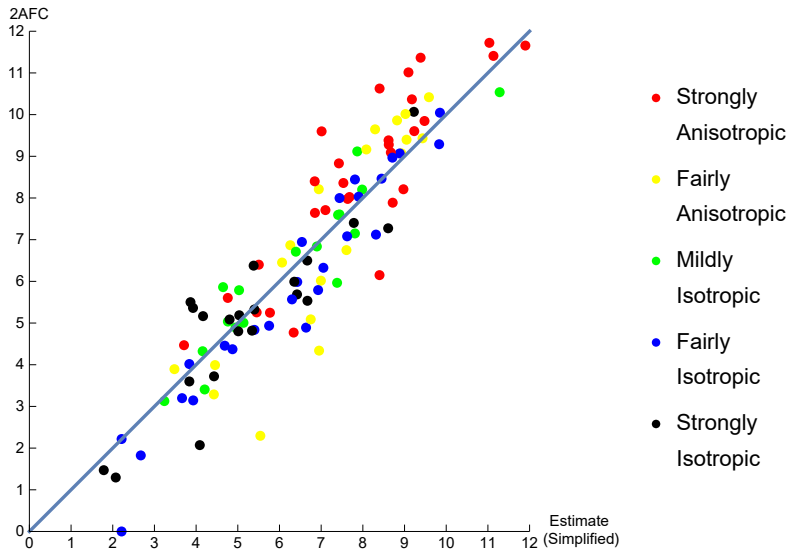


Figure 4.24: Algorithm estimates *vs.* 2AFC coloured according to isotropy.

We also evaluated the performance of the algorithm using the SURF curves introduced in Section 4.1.2.4. The four $\log_2$ keypoint curves were mapped to the perceptual ground truth texture scale estimates arising from the 2AFC task.

For comparison, we have implemented the methods of Ardizzone et al. (2013), and Hong et al. (2008) reviewed in Chapter 2. The method of Hong et al. (2008) gives estimates on a per-pixel basis. Given that the Brodatz set textures each have

uniform scale, we take the mean overall local scale estimates in an image as a measure of global texture scale.

The scope of the Ardizzone et al. (2013) algorithm is narrower than ours, as it is only intended for regular, near-regular and homogenous textures. So, to allow for a fair comparison, we also report results for a subset of the Brodatz dataset. The subset, selected by the authors, comprising 24 regular, near-regular and homogeneous textures with not overly large or small repeating elements from the Brodatz dataset, was chosen to match the scope of the method of Ardizzone et al. (2013) based on the information provided in their work. The textures in this set are indicated by an asterisk in Table 4.2.

	All Brodatz	Regular Subset
Baseline	2.5 [2.3 - 2.8]	1.6 [1.3 - 2.0]
Ardizzone et al. (2013) (*n* keypoints)	2.5 [2.2 - 2.8]	1.5 [1.1 - 1.9]
Ardizzone et al. (2013) (keypoint strength)	2.5 [2.2 - 2.8]	1.5 [1.1 - 2.0]
Hong et al. (2008)	1.7 [1.4 - 1.9]	1.1 [0.8 - 1.3]
This work - SURF curves	2.3 [2.0 - 2.6]	1.3 [1.1 - 1.6]
This work - simplified	1.4 [1.2 - 1.6]	1.0 [0.7 - 1.3]
This work - all curves regression	**1.2** [1.0 - 1.4]	**0.6** [0.3 - 1.0]

Table 4.1: RMS errors with 95% confidence intervals, measured in jnds, between algorithm-predicted and 2AFC-estimated values on the perceptual interval scale.

Table 4.1 shows the RMSE between predictions and the 2AFC dataset for the different methods assessed and for a baseline method. The left column labelled 'Repeating Subset' comprises the 24 aforementioned textures selected by the authors. The first row, labelled Baseline, displays the RMSE obtained by predicting the scale of each texture by the population mean. This is the optimum algorithm that provides texture scale predictions without considering the texture. The second and third rows of Table 4.1 show the RMSE for the method of Ardizzone et al. (2013), using a fixed number of keypoints, and a fixed keypoint strength across images respectively. We note that these algorithms fail to perform significantly above the baseline for the full dataset and the repeating subset. The fourth row shows the performance obtained using the SURF keypoint density curves. The performance is better than baseline, and that or Ardizzone et al. (2013), but is behind the Hong et al. (2008), and the

out algorithm based on BIF curves. The fifth row shows the RMSE for estimates obtained using the method of Hong et al. (2008). The performance of this algorithm is consistently better than baseline, and close to within human sensitivity for the repeating subset. The sixth row shows that the performance of our simplified approach is better than that of Hong et al. (2008), but not significantly, as there is an overlap of the confidence intervals. The performance of this simplified method validates our hypothesis that texture scale is determined by the increase in similarity of the appearance of regions, as the region size increases, which is crudely captured by this univariate feature. The bottom row shows the performance of our method when using all the points on all the texture curves as features for the regularised linear regression. The greatest improvement (0.2*jnds*) of our full method over our simplified method is due to the regression extracting information from the variation of the 25 curves, and/or from the shape of their fall, not just their initial slope. Our full method performs significantly better than previous methods. On the full Brodatz dataset it does not quite achieve an error of 1.0, which would signal effective agreement with human perception, but on the repeating subset it does.

In order to assess the agreement between the perceptually derived estimates, the algorithm derived estimates, and real measurable scales (or texel-based texture scale estimates), we report the pairwise correlations between these in Table 4.3. The final column, 'Sqrt Area', is only assessed for the 24 repeating Brodatz textures. These 24 textures are indicated in Table 4.2. The cropped texels are shown in Figure 4.25. We note that there is a high correlation between our perceptual estimates and square-root area, and higher agreement between square-root area and the estimates derived using our algorithm than competing algorithms.

4.5 Summary and Discussion

Having obtained ground-truth data against which to validate candidate models we proposed an algorithm for estimating texture scale. To do so we first computed the BIF representation of a texture at various BIF parameters, then we measure how the variance across the position of the BIF histograms of local regions changes with

Tex	Scale (jnd)	Scale (px.)	Tex	Scale (jnd)	Scale (px.)	Tex	Scale (jnd)	Scale (px.)	Tex	Scale (jnd)	Scale (px.)
D1^{R*}	5.99	90	D29^{I}	3.14	60	D57^{I}	3.20	61	D85^{N}	5.60	85
D2^{I}	7.12	105	D30^{H*}	9.29	143	D58^{I}	11.01	182	D86^{I}	7.61	113
D3^{N}	4.94	88	D31^{H*}	10.05	157	D59^{I}	11.37	192	D87^{I}	6.45	96
D4^{I}	3.60	64	D32^{I}	2.07	52	D60^{I}	8.20	123	D88^{H*}	10.54	171
D5^{I}	6.50	97	D33^{I}	4.82	76	D61^{I}	9.65	150	D89^{H*}	9.40	145
D6^{R}	1.82	50	D34^{R}	4.37	72	D62^{I}	10.37	166	D90^{I}	10.01	158
D7^{I}	10.07	160	D35^{N}	5.79	87	D63^{I}	9.12	140	D91^{I}	11.66	200
D8^{N}	6.75	100	D36^{N}	5.01	78	D64^{R*}	6.94	103	D92^{I}	5.36	82
D9^{I}	3.99	68	D37^{I}	7.98	119	D65^{R*}	5.86	88	D93^{I}	5.26	81
D10^{N*}	8.36	125	D38^{I}	4.77	76	D66^{H}	5.99	90	D94^{N*}	7.08	105
D11^{N}	6.87	102	D39^{I}	9.60	150	D67^{I}	6.15	92	D95^{N*}	6.33	94
D12^{I}	7.15	106	D40^{I}	9.85	155	D68^{I}	5.25	81	D96^{N*}	9.09	139
D13^{I}	9.07	139	D41^{I}	9.60	149	D69^{I}	8.83	134	D97^{I}	9.16	141
D14^{R}	3.13	60	D42^{I}	10.63	173	D70^{I}	8.40	126	D98^{H}	8.00	119
D15^{I}	7.64	113	D43^{I}	11.72	201	D71^{I}	9.38	145	D99^{H}	9.86	155
D16^{R}	0.00	39	D44^{I}	11.41	193	D72^{I}	9.28	143	D100^{I}	6.71	100
D17^{R}	3.89	67	D45^{I}	10.42	168	D73^{I}	5.97	90	D101^{R*}	5.19	80
D18^{N*}	7.59	113	D46^{R*}	8.02	120	D74^{H*}	8.03	120	D102^{R*}	5.33	82
D19^{N}	4.89	77	D47^{R*}	8.21	123	D75^{H*}	7.27	108	D103^{N}	4.80	76
D20^{R*}	5.50	84	D48^{R*}	9.07	139	D76^{I}	5.04	79	D104^{N}	5.09	79
D21^{R}	1.29	46	D49^{R}	4.47	73	D77^{R}	1.47	48	D105^{I}	5.09	79
D22^{N}	4.84	76	D50^{I}	7.71	115	D78^{N}	2.30	53	D106^{N}	4.34	71
D23^{H}	8.44	127	D51^{I}	7.89	117	D79^{N}	3.29	61	D107^{I}	8.46	127
D24^{I}	4.32	71	D52^{R}	4.02	68	D80^{N}	4.46	72	D108^{I}	8.97	137
D25^{R*}	9.43	146	D53^{R}	2.22	53	D81^{I}	4.89	77	D109^{I}	5.79	87
D26^{R*}	8.21	123	D54^{H}	5.53	84	D82^{N}	5.17	80	D110^{I}	5.57	85
D27^{H}	7.40	110	D55^{N}	3.41	62	D83^{R*}	6.40	95	D111^{I}	6.38	95
D28^{H}	5.69	86	D56^{R*}	6.02	90	D84^{I}	3.72	65	D112^{I}	6.84	101

Table 4.2: Texture scale ground-truth values (in jnds and pixels) for the 112 Brodatz textures (D1-D112), as estimated from the 2AFC experiment. Superscript denotes regularity as per our taxonomy (R: regular; N: near-regular; H: homogeneous, I: irregular). Asterisk indicates textures included in the repeating subset in Table 4.1.

window size, and finally we transformed these variances into texture scale estimates with the help of a regularised linear regression.

Our method makes texture scale predictions with an RMS error, compared to human ground-truth, of 1.2*jnds*. The error of our method is still more than the error of 1.0*jnds* that would be considered a practical match, but it is good enough for some applications and it is the only method that calculates texture scale for non-regular texture to below 1.7*jnds* (Hong et al., 2008). Furthermore, we find that it performs better than methods that work on regular, near-regular and homogeneous textures, even in datasets that should favour the latter. Our method performs well on textures of all regularities and both isotropic and anisotropic textures.

The estimates produced by our method are highly correlated to both perceptual

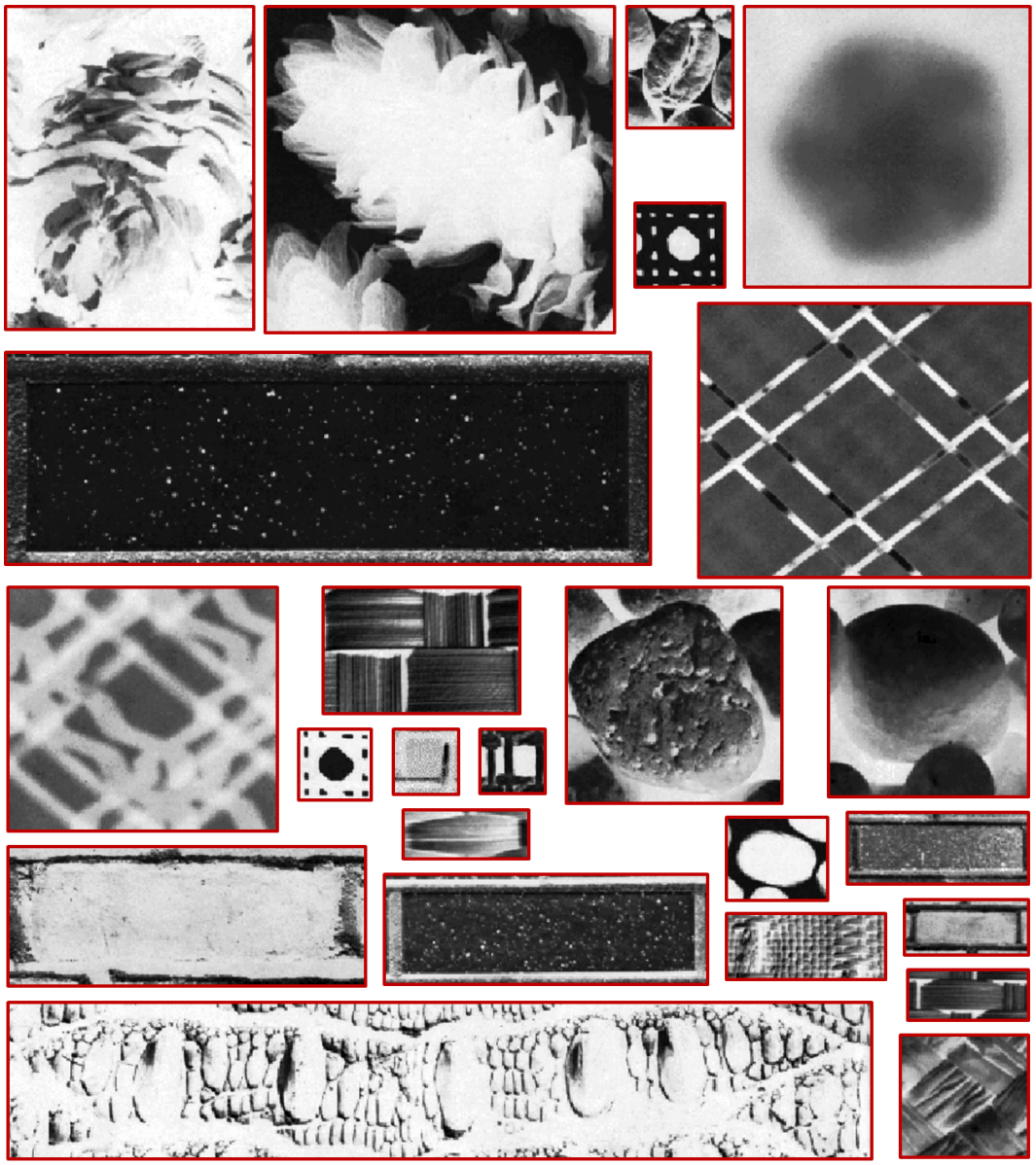


Figure 4.25: 24 Texels cropped from Brodatz textures with a clearly discernible repeating element as indicated in Table 4.2.

derived datasets and to the size of manually segmented texels for textures with a clearly discernible repeating unit. This indicates that despite producing estimates in jnd units, a linear transformation followed by an exponentiation enables us to get estimates in pixels which are highly consistent with physically measurable scales.

We also proposed a simplified algorithm which is $10x$ less computationally expensive yet $70-80\%$ as accurate, and still more accurate than the much more computationally expensive algorithms by Ardizzone et al. (2013) and Hong et al.

	Absolute Scaling	2AFC	This work - SURF curves	This work - full	This work - quick	Ardizzone et al. (2013)	Hong et al. (2008)	Sqrt Area (Reg)
Absolute Scaling	-	0.90	0.44	0.86	0.82	0.40	0.71	0.89
2AFC		-	0.44	0.92	0.86	0.39	0.77	0.88
This work - SURF curves			-	0.44	0.48	0.72	0.27	0.40
This work - full				-	0.93	0.37	0.81	0.76
This work - simplified					-	0.48	0.78	0.66
Ardizzone et al. (2013)						-	0.32	0.17
Hong et al. (2008)							-	0.60
Sqrt Area (Reg)								-

Table 4.3: Pairwise correlations between the two perceptually derived datasets (Absolute Scaling and 2-AFC); the estimates derived using the algorithm proposed in this paper; the estimates derived using the methods by Ardizzone et al. (2013) and Hong et al. (2008); and square-root texel area (manually segmented by author) for 24 regular and near-regular Brodatz.

(2008).

One limitation of our approach is that the performance of the algorithm is highly sensitive to the choice of the two parameters, σ_B and γ, controlling the BIF representation. We have chosen a set of BIF parameters that gives the best performance for images of a size similar to the Brodatz (i.e. 640×640 pixels). Magnifying an image significantly, say four fold, would require us to adjust the BIF parameter settings to ensure that the BIF representation captures the structure of images of that size. This in itself is a scale-selection exercise that we deemed to be beyond the scope of this work, and we suggest it as an avenue for further research in Chapter 6.

Chapter 5

An Online Tool for Calculating Texture Scale

In Chapter 3 we showed that the human visual system is capable of assessing texture scale consistently for textures of all regularities. In Chapter 4 we presented an algorithm for the quantification of texture scale and showed that the estimates produced by our algorithm are consistent with human assessment. Having done this we thought that this work could be useful to researchers across the spectrum of scientific disciplines, so we wanted to *(i)* make it available to researchers, and *(ii)* obtain feedback from researchers on our work. The purpose of the feedback was to determine whether the estimates produced by our algorithm are in line with researchers' expectations and why. In order to make our algorithm accessible, we created an online tool (accessible via `http://www.texturescalcalculator.com`) to which users are able to upload images of a texture, and obtain an estimate of the scale for the uploaded texture.

5.1 The Interface

Figure 5.1 shows the landing page of our interface. It includes our characterisation of texture scale and an image upload facility. Users have the option to either drag and drop an image from their computer into the upload box, or to upload images using a standard browse box (similar to the process for adding an attachment to an email). The image needs to be in a standard picture format (e.g. JPG, PNG,

BPM, ...), no smaller than 256×256 pixels and no larger than 4000×4000 pixels. For optimal performance, we recommend uploading images larger than 512×512 pixels and smaller than 1024×1024 pixels. Our algorithm is designed to work for images showing constant textures (i.e. only one type of texture which is fairly consistent across the image), with no background regions and few image defects (e.g. luminance features, perspective skew, ...).

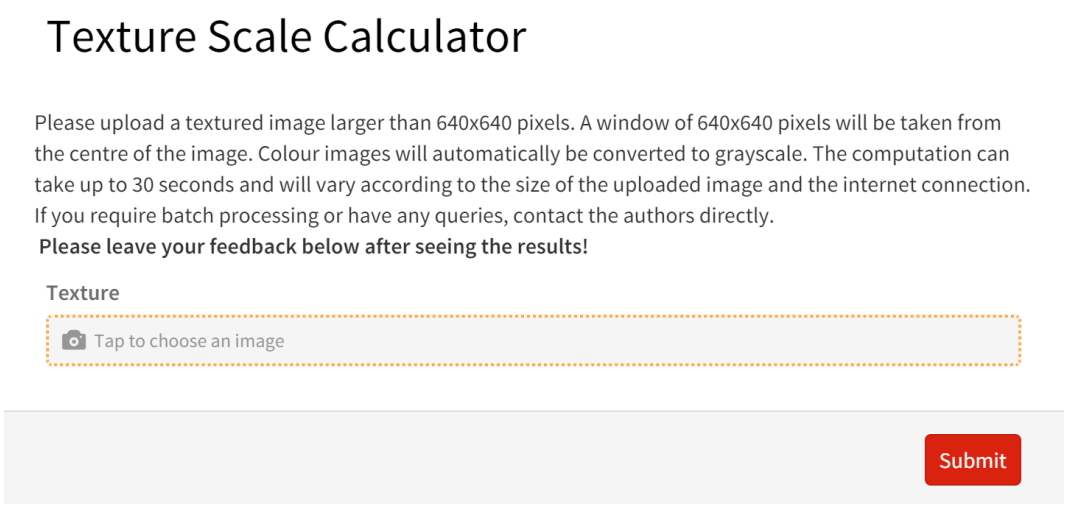


Figure 5.1: Landing page for online calculator.

Once the image is uploaded, users can click on the "Submit" button. The submitted image undergoes initial pre-processing. First, we check that the image satisfies the size and format requirements, then it is converted to gray-scale (if the original image was coloured), and resized to 640×640 pixels. We resize the images because the BIF parameters used in our algorithm are optimised for images of this size. After the pre-processing stages, the image is passed through our texture scale algorithm. The algorithm is implemented using Wolfram Mathematica and runs on the Wolfram Cloud platform. It takes approximately 20 seconds to get a texture scale estimate, but can take up to 3 minutes depending on the size of the uploaded image and internet connection speed. The scale is estimated using our "simplified" algorithm because of computational tractability. Providing estimates using the complete method would take 5-15 minutes per image to compute given the single-thread processing restriction of Wolfram Cloud. The interface then outputs the cropped /

resized gray-scale texture with an upright square window overlaid. The size of the window corresponds to the texture scale and its position on the image is arbitrary. We use a square window to show texture scale because our algorithm produces estimates in jnd units. We then use the conversion rule presented in Section 3.4 to obtain an estimate in pixels. Recall that our empirical conversion rule was obtained by comparing the perceptual texture scale estimates derived from our two psychophysics studies, and that the estimates from the first experiment (absolute scaling) were based on upright square windows. It would therefore be misleading to show anything other than upright squares when using this conversion. Figure 5.2 shows the output of our interface for an example texture.

For confidentiality reasons, we do not ask users to identify themselves and we do not store the images uploaded to the interface. A group of experts and prospective users were invited to try the interface in order to obtain some feedback. We discuss the collected feedback in the next section (5.2).

5.2 User Feedback

In order to assess whether the estimates produced by our algorithm are in line with end-user expectations, we carried out two rounds of user feedback on the output. The first round was done through the online interface, and the estimates produced using the simplified version of our algorithm. The second round,included A-B testing, in which subjects were sometimes presented with either much larger or smaller estimates than our algorithm would predict in order to ensure that volunteers were not being overly positive in their feedback, and the estimates were produced using the complete version of the algorithm. The two rounds are described in the sections 5.2.1 and 5.2.2 respectively.

5.2.1 First User Test

We collected feedback from a cohort of experts. The experts comprised PhD students and Post-Doctoral students at UCL, as well as people working on image analysis products in the industry. PhD students from CoMPLEX, which covers a wide range of scientific disciplines, were also part of the cohort. In total, 54 people were

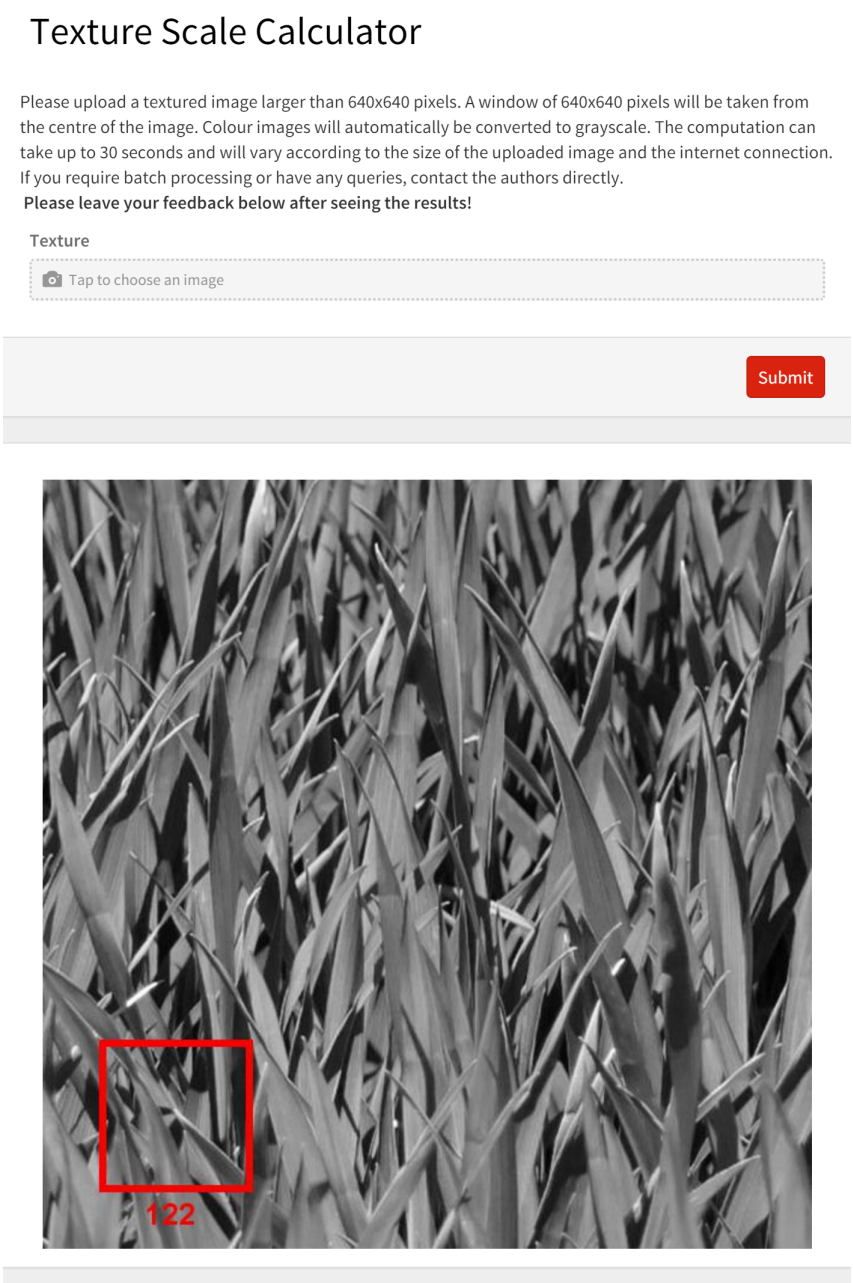


Figure 5.2: Interface output showing the cropped gray-scale converted texture and an overlaid window of size corresponding to the texture scale.

invited to take part in the feedback round. We collected 65 feedback responses from 19 users over a month-long period. Figure 5.3 shows the feedback form. Respondents were asked, "Given our characterisation of texture scale, how does the indicated scale seem to you?". They could select one of five possible options: "Much too small", "A bit too small", "About right", "A bit too large", or "Much too large".

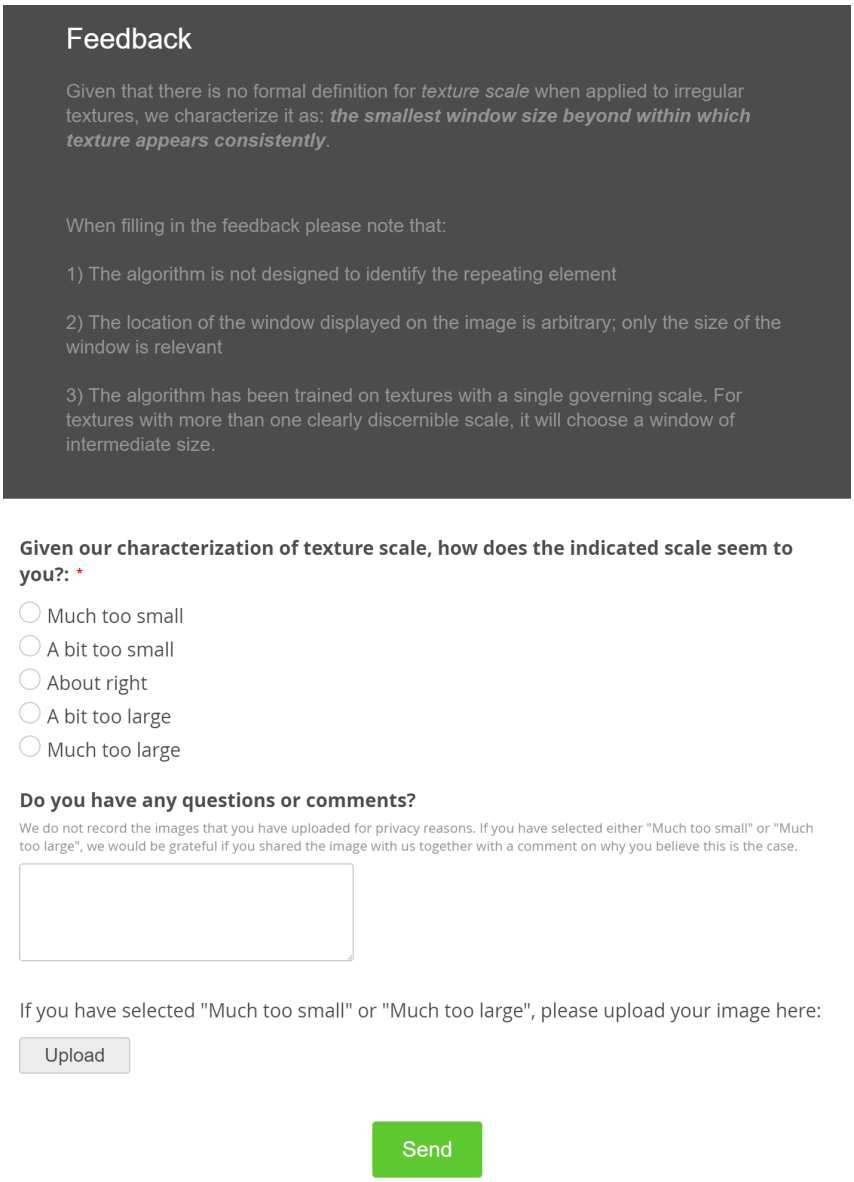


Figure 5.3: Feedback form shown below the interface output.

Respondents were then allowed to leave questions and comments in free-text form and invited to share with us textures for which they considered the estimates to be "Much too small" or "Much too large".

Figure 5.4 shows a histogram of the collected feedback responses. There were no "Much too large" responses, indicating that our algorithm does not often majorly overestimate texture scale. There were four (out of 65, $\approx 6\%$) "Much too small" responses. Three out of these four textures were shared with us and are shown in Figure 5.5. The left image in Figure 5.5 shows a concrete brick pavement. The scale

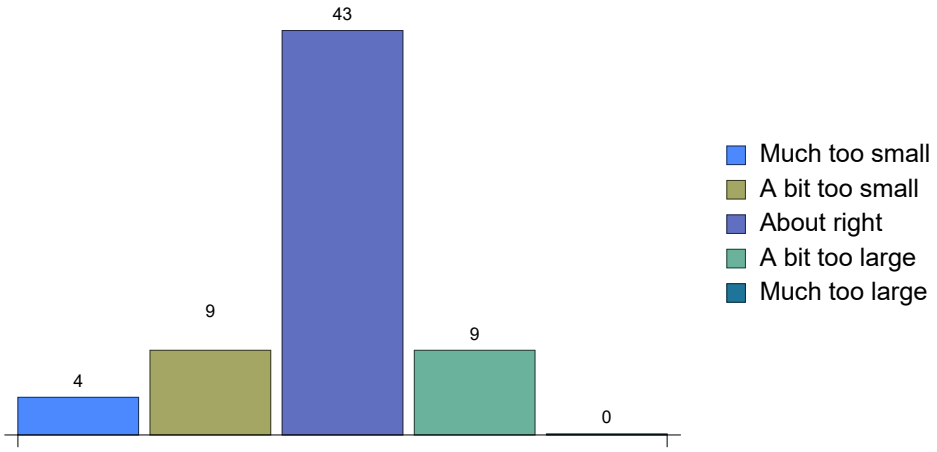


Figure 5.4: Histogram of feedback responses.

estimate is much too small as it is does not show a full slab. This is an isotropic texture with strong perspective skew with which our algorithm could have struggled. In addition, there is texture within the individual bricks (concrete) and our algorithm could have picked up variation at the concrete scale instead of the brick scale. The centre image in Figure 5.5 was submitted to us with a comment saying, "I would have expected a window of at least 6 x 6 circles to capture the 3 x 3 tile pattern variation". Our algorithm was not trained on textures displaying regularity at more than one scale, so we would not expect the algorithm to cope with this type of texture. The algorithm appears to have selected an intermediate size window; one that is a bit too large for the dot-scale regularity but much too small to capture regularity at the 3x3 dot scale. The image on the right of Figure 5.5 shows a piece of furniture with two very different regions that are highly magnified and do not constitute a uniform texture, so we are not surprised to see that our algorithm was not able to cope with such a region.

5.2.2 Second User Test

In order to ensure that respondents were not being overly generous with the feedback, we carried out a second round of end-user feedback. In the second round, a cohort of 15 people were asked to send the author 3-10 images of textured surfaces. We received 33 images from 10 volunteers. The images were converted to grayscale, cropped to a square shape and resized to 640×640 pixels.

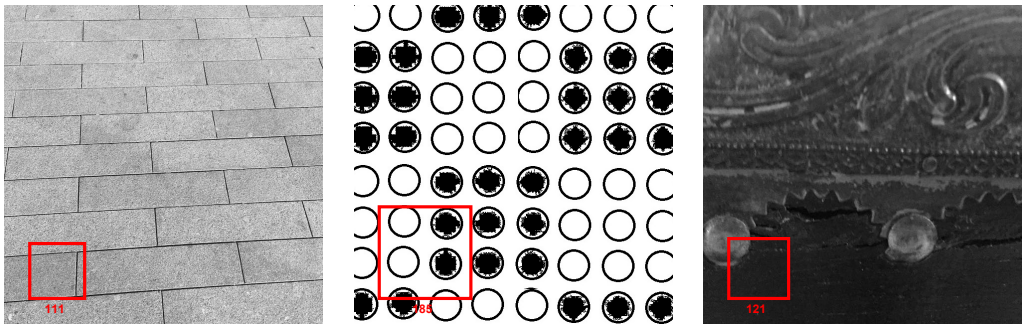


Figure 5.5: Three textures for which the estimates of our simplified algorithm were deemed to be "much too small" by test users.

We calculated the texture scale for each as per our complete algorithm, and recreated the same output as for the first round of feedback, i.e. a red box superimposed on the image. We produced three versions of the output:

- Texture scale estimate half the size of that predicted by our algorithm
- Correct size texture scale estimate
- Texture scale estimate twice the size of that predicted by our algorithm

Subjects were sent the processed versions images they submitted with one of the three texture scales estimates mentioned above. In addition, they were some additional images submitted by other volunteers in order to boost the number of responses. For each image, they were asked to submit feedback on the window size, in the same manner as per the interface feedback form. We collected a total of 97 responses. Figure 5.6 shows the histogram of feedback on estimates half the size of those from our algorithm (left), the correct size (middle), and twice the size of those from our algorithm (right). The feedback contained more "a bit too small" responses when for the internationally smaller windows, and "a bit too large" for intentionally larger windows.

When participants were shown estimates half the size of those predicted by our algorithm, most of the feedback was that the estimates were too small (Figure 5.6, left). When participants were shown estimates predicted by our algorithm, most of the feedback was that the estimates were as expected (Figure 5.6, middle). When participants were shown estimates twice the size of those predicted by our

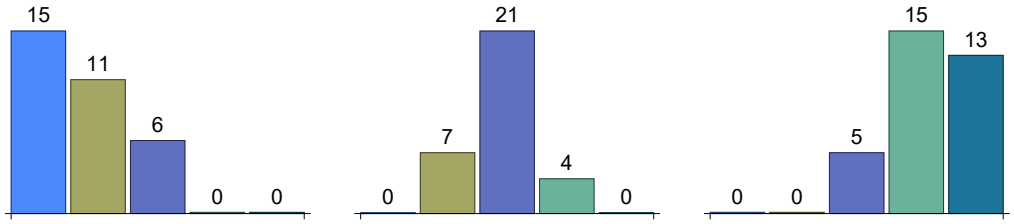


Figure 5.6: Histogram of feedback on estimates half the size of those from our algorithm (left), the correct size (middle), and twice the size of those from our algorithm (right). Columns are "much to small", "a bit too small", "about right", "a bit too large" and "much too large", left to right in each bar-chart.

algorithm, most of the feedback was that the estimates were too large (Figure 5.6, right). This feedback is consistent with that from the first feedback round, in which there was no A-B alternative.

5.3 Summary

In Chapter 3 we showed that human observers are able to assess texture scale consistently, irrespective of regularity, and in Chapter 4 we developed an algorithm for estimating the scale of texture. In this chapter we presented an online tool that we have made available to the wider public. Users can upload images of textures to our online tool and obtain estimates of the scale of the uploaded textures. The estimates are obtained using the simplified version of our algorithm (as per Section 4.3.2). While the complete algorithm presented in Section 4.3.1 is more accurate, the single thread computational restriction of the platform on which the tool is hosted (Wolfram Cloud) would take several minutes to produce an estimate using this method.

We collected feedback from a 17-strong cohort reported that the estimates produced by our algorithm were in line with their expectations approximately 80% of the time, a bit too large or a bit too small approximately 15% of the time, and much too small approximately 5% of the time. The cohort comprised PhD and post-doctoral students at UCL CoMPLEX covering a wide range of scientific disciplines, and image analysis developers working in start-ups in London. We are very satisfied with the feedback given that it was collected on non-controlled images. Furthermore, the estimates were obtained using our simplified algorithm, translated from jnd units to pixels, and illustrated by upright square windows. We have iden-

tified that textures with more than one level of regularity are problematic for our algorithm as these textures have multiple scales. We will provide a more detailed critical review of our algorithm's strengths and opportunities for further development in Chapter 6. An additional limitation of the online calculator is inherited from the algorithm itself - the BIF parameter settings have been fixed and optimised for images of a similar size to the Brodatz (i.e. 640×640 pixels). As a result, if an image larger than this size is submitted, only the central portion of the image is assessed. This limitation is discussed in more detail in the critical evaluation provided in Chapter 6.

Chapter 6

Summary & Conclusions

Scale is a highly important topic in the image analysis and vision literature which we discussed at length in Chapter 2. The vast majority of the discussion on texture scale in the texture analysis was under the paradigm that texture was formed by a repeating element (texel) and a placement rule governing how the texel is placed on the scene. Under this paradigm, texture scale was defined as the size of the texel, distance between texels, window size at which the distribution of the number of keypoints was most uniform, and as the smallest window size at which a function of domain similarity dipped (amongst many other definitions). While these definitions are good enough for regular, near-regular and homogeneous textures, all of which are formed by the placement of a repeating element on the scene, they are ill-defined for irregular/stochastic textures lacking a clearly defined repeating element. Does this mean that texture scale ought to be undefined for such irregular/stochastic textures? No, we observed that irrespective of whether a texture is regular or not, samples of increasing size taken from a texture appeared increasingly similar. There must therefore be a smallest window size beyond within which samples taken from a texture appear consistent. We thus put forward the hypothesis that the concept of scale ought to extend to textures of all regularities. In order to test this hypothesis we needed a way to characterise texture scale; however, there was no mathematically rigorous definition in the literature that extended to irregular/stochastic textures. Instead of devising an arbitrary, yet rigorous, definition, we proposed a more relaxed perceptual characterisation of texture scale: *the smallest*

window size beyond within which texture appears consistently.

Having characterised texture scale, we set ourselves three objectives for this work:

1. To check whether humans were able to assess texture scale consistently as per our characterisation, and if so, measure perceptual assessment of texture scale for a gold-standard texture database;

2. To develop an algorithm for estimating the scale of textures that *(i)* was validated against human assessment, and *(ii)* worked for all textures, those which are formed by a clearly discernible repeating element, and those which are not; and

3. To make our research available to the wider scientific community for use, critique and improvement.

We are pleased to have achieved the above objectives. By means of two perceptual experiments, we confirmed that human observers are able to assess the scale of texture consistently, irrespective of texture regularity. Through these two experiments, we collected perceptual ground-truth texture scale values for the Brodatz dataset. We then proposed an empirical algorithm capable of estimating texture scale with an RMS error of 1.2jnds compared to the ground-truth for textures of all regularities, and an RMS error of 0.6 jnds compared to the ground-truth for textures with a clearly discernible repeating element. Both of these scores are significantly higher than any other competing methods in the current literature. Our algorithm is both *(i)* validated against human assessment, and *(ii)* able to handle textures of all regularities. The error of our method is still more than the error of 1.0jnds that would be considered a practical match, but it is good enough for some applications and it is the only method that calculates texture scale for non-regular textures to below 1.7jnds (Hong et al., 2008). Finally, we developed an online calculator to which the wider public can upload images of textures and obtain texture scale estimates as per our algorithm.

In Section 6.1 we provide a critical appraisal of the contributions contained in this thesis. In Section 6.2 we suggest candidate avenues for the further development of this work.

6.1 Critical Appraisal of Contribution

The five primary contributions presented in this thesis are the following:

A new characterisation of texture scale: We propose a new characterisation of texture scale that is devoid of any assumption with regards to the structure or definition of texture, and that is applicable to all textures, irrespective of whether they are formed by a repeating element placed on the scene according to a regular grid or not.

Evidence that human observers are able to assess texture scale irrespective of regularity: We show that humans are able to assess texture scale irrespective of its regularity. In particular, we show that human assessment is consistent within and between subjects, and across methods of probing.

Perceptually validated ground-truth texture scale values for the Brodatz dataset: We measure human assessment of texture scale for the Brodatz dataset and provide estimates of the variability of each of the Brodatz. We make this dataset publicly available so as to enable researchers to make use of it as ground-truth in the validation of other candidate methods and applications.

Image analysis algorithm for estimating texture scale: We develop an algorithm for estimating texture scale which is validated against human perception, and show that experts find the output of our algorithm useful on images from a wide range of disciplines.

Online tool for estimating texture scale: We make our research available by means of an online calculator to which researchers and the wider public can upload images of textures and obtain estimates of their texture scale.

In the next sections we provide a brief summary of the contributions of this work as well as a critical evaluation for each contribution.

6.1.1 A New Characterisation of Texture Scale

Scale, when referring to texture, has been defined in numerous ways. It has been defined as the size of a repeating unit, or equivalently, as the mean distance between the repeating units (Malik et al., 2001). Ardizzone et al. (2013) define the scale of texture as the smallest upright square window containing a consistent number of image keypoints. Hong et al. (2008) define scale as the smallest upright square window for which tiled neighbouring windows of the same size are similar. As mentioned earlier in this chapter, these definitions only work under the paradigm of texture being formed by the placement of a clearly discernible repeating element. Our characterisation, *the smallest window size beyond within which texture appears consistently*, has several advantages over the aforementioned, more rigorous, definitions. First, by virtue of being a perceptual characterisation, it is agnostic to the paradigm of choice for texture. Second, it is well-defined for texture spanning the entire regularity spectrum, from regular to irregular/stochastic textures. Third, it is flexible, in the sense that it does not impose a measure for 'consistency' - this is left up to the researcher. The main limitation of such a characterisation, true for many other generic characterisations, is that there is no single correct way of measuring texture scale as per its wording. We tried to remain as true to our perceptual characterisation as possible by opting for an empirical algorithm.

6.1.2 Evidence of Human Ability to Assess Texture Scale Irrespective of Regularity

Having coined how we wanted to characterise texture scale and having observed that samples of increasing size taken from a texture appear increasingly consistent as we made the sample window larger, we wanted to check whether the human visual system was able to make sufficiently consistent assessments (between and within subjects, and across methods of probing) so as to enable us to measure human assessment of texture scale for a gold-standard texture database, the Brodatz library.

We carried out a first direct-scaling experiment and found consistency between and within subjects in the perception of texture scale on the Brodatz library. This

first experiment suffered from two main limitations. First, there appeared to be some systematic subject bias with some subjects selecting consistently smaller or larger windows. This type of bias is common amongst direct-scaling methods so we performed a procrustes analysis enabling us to correct for some of this effect. Second, despite the procrusted correction, the standard deviations for some of the estimates were a bit larger than we would wish them to be, we suspect due to the small number of subjects ($n = 8$).

We carried out a second indirect-scaling experiment (2AFC type), devoid of the systematic bias issue, with a much larger number of observers ($n = 40$) which enabled us to obtain perceptual ground-truth texture scale values for all the Brodatz on an interval scale. The interval scale spanned 12jnds, indicating that subjects were able to ditinguish 12 levels of texture scale within the Brodatz database. We were pleased to see a correlation of 0.9 between the log window sizes (in pixels) obtained in the first experiment and the 2AFC derived interval scale values. This is consistent with Fechner (1966) and Gescheider (1997) who suggest that if increment thresholds obeyed Webber's Law (which states that increment magnitude detection thresholds are proportional to stimulus magnitude), the underlying psychophysical scale can be approximated by a logarithmic transform. The strong correlation between the jnd and pixel unit scales enabled us to convert jnd estimates produced by our algorithm into estimates in pixel units.

6.1.3 Perceptually Validated Texture Scale Values for the Brodatz

We made available, in Chapter 3, the texture scale estimates for all Brodatz textures obtained through each of these two experiments, as well as standard deviations for each. While this is the only perceptually validated dataset for texture scales available in the literature, there is scope to extend it to other texture libraries and types of image.

6.1.4 Algorithm for Estimating Texture Scale

We proposed an empirical image analysis algorithm for the estimation of texture scale. The algorithm estimates texture scale according to how the variance of BIF histograms changes with window size. Our algorithm is able to estimate the scale for the Brodatz with an RMSE of 1.2 jnds, and 0.6 jnds for textures with a clearly discernible repeating unit. This performance is far superior to anything available in the literature. The estimates obtained by our algorithm are also highly correlated with the size of texels (manually segmented) for textures with a clearly discernible repeating element. In addition to the complete algorithm, we proposed a simplified version which is $70-80\%$ as accurate as the complete algorithm but 10 times less computationally expensive. These performance estimates are based on the code being implemented in Mathematica. Further performance improvements are likely to be realised if translated to lower level programming languages. To the best of our knowledge, this is the only algorithm in the literature that is designed to estimate the scale of textures of all regularities. It is also the best performing algorithm for regular textures when compared to Hong et al. (2008) and Ardizzone et al. (2013), the two most similar methods in scope from the literature. While the performance of our algorithm is superior to its competitors, it suffers from two main limitations. The first limitation is that its accuracy is reliant on the BIF representation capturing the structure of the texture effectively. To accomplish this, the BIF parameters must be carefully tuned. The consequence of this is that our algorithm is not zoom-invariant. We selected the BIF parameters in our algorithm for optimal performance on images the size of the Brodatz (640×640 pixels). The second limitation is that textures often display regularity at multiple scales, and thus the scale of a texture may not be captured by a single number. We propose two directions of further research addressing these two limitations in Section 6.2.

6.1.5 Online Texture Scale Calculator

Very often, research carried out in the image analysis field never sees practical use, beyond publication. Because this research was funded through an interdisciplinary doctoral training centre, we wanted to make our research available to the wider aca-

demic community for use, critique and improvement. We have therefore created an online tool to which users can upload images and obtain estimates of texture scale as per our algorithm. Unfortunately, due to the computational restrictions of the platform on which it is hosted, we were restricted to implementing the simplified version of our algorithm. In addition, the interface does not allow for batch computing, so textures need to be uploaded one at a time.

6.2 Suggested Direction of Further Research

This work has been submitted for publication to the Image, Vision and Computing Journal and is currently under second round of review - corrections were submitted to the journal on Jan 9th 2018.

The performance of our algorithm is sensitive to the choice of BIF parameters. The BIF parameters used in this work were chosen so as to ensure that they capture structure at all possible relevant scales present in images the size of the Brodatz (640x640 pixels). The procedure we adopt when estimating texture scale for images of different sizes, is to rescale the test image to 640x640 pixels in such a way that the structure of interest is clearly discernible at this resolution. Doing so leads to good quality estimates overall, as indicated by the A-B feedback we collected on images submitted by prospective users (Section 5.2). We would therefore discourage prospective users from adjusting the parameters to match the image size, and instead encourage them to rescale their images, as this is quicker, easier, and the results have been validated.

Textured images may have structure at multiple scales. This cannot be summarized by a single number. We have made a well-defined and validated step towards quantifying the scale of texture which performs well on images exhibiting a single level of structure. 640x640 pixels is generally a large enough area to capture variation at a single scale.

A possible next step in this line of research would be to develop a vector representation of texture scale that would measure how strong the scale signal is at a range of scales. Consider an image of a brick wall; there could be structure at the

brick level, and at the brick-material level. A 640x640 pixel image could struggle to capture structure at both the brick and material level. An algorithm looking to successfully capture multiple scales within an image would, in all likelihood, need to be able to handle images larger than 640x640 pixels.

A second consequence of fixed BIF parameters, is that our the performance of our algorithm may be affected under magnification. To achieve magnification-invariance, the BIF parameters need to be selected automatically for the image being processed.

Another possible avenue of further research is the use of anisotropic windows. Our algorithm uses isotropic soft-Gaussian windows which may not capture repeating elements of anisotropic textures.

Colophon

- This document was set in the Times Roman typeface using LaTeX and BibTeX, composed with a text editor.
- Zotero was used as a reference managing tool.
- The software development and implementations were done using Wolfram Mathematica.
- The figures and graphs were produced using Wolfram Mathematica and edited with Gimp, Inkscape and Windows Paint.
- The interface for the direct scaling experiment presented in Section 3.2 was implemented using JavaScript, as well as the JQuery, JCrop and Ajax packages.
- The front-end of the interface for the second experiment, presented in section 3.3 was developed on JavaScript, and the back-end was managed using Stamplay, which enable the responses to be stored in an easy to access Google Sheet document.
- The texture scale calculator was developed using the Wolfram Mathematica Cloud computing and deployment features. The website was built with Wix. The feedback form was created using a tool provided by 123ContactForm.

Bibliography

Abdelmounaime, S., Dong-Chen, H., Feb. 2013. New Brodatz-Based Image Databases for Grayscale Color and Multiband Texture Analysis. International Scholarly Research Notices 2013, e876386.

Acharya, U. R., Ng, E. Y. K., Tan, J.-H., Sree, S. V., Ng, K.-H., 2012. An integrated index for the identification of diabetic retinopathy stages using texture parameters. Journal of medical systems 36 (3), 2011–2020.

Andrews, J. T. A., Tanay, T., Edward, M. J., Griffin, L. D., 2016. Transfer representation-learning for anomaly detection. Vol. 48. W&CP, New York, NY, USA.

Ardizzone, E., Bruno, A., Mazzola, G., 2013. Scale detection via keypoint density maps in regular or near-regular textures. Pattern Recognition Letters 34 (16), 2071–2078.

Arivazhagan, S., Ganesan, L., 2003. Texture classification using wavelet transform. Pattern recognition letters 24 (9), 1513–1521.

Arivazhagan, S., Ganesan, L., Priyal, S. P., 2006. Texture classification using Gabor wavelets based rotation invariant features. Pattern recognition letters 27 (16), 1976–1982.

Arvis, V., Debain, C., Berducat, M., Benassi, A., 2011. Generalization of the cooccurrence matrix for colour images: application to colour texture classification. Image Analysis & Stereology 23 (1), 63–72.

Backes, A. R., Casanova, D., Bruno, O. M., May 2012. Color texture analysis based on fractal descriptors. Pattern Recognition 45 (5), 1984–1992, wOS:000301017700015.

Baştürk, A., Günay, E., 2009. Efficient edge detection in digital images using a cellular neural network optimized by differential evolution algorithm. Expert Systems with Applications 36 (2), 2645–2650.

Bay, H., Tuytelaars, T., Van Gool, L., 2006. Surf: Speeded up robust features. In: Computer vision–ECCV 2006. Springer, pp. 404–417.

Bhattacharyya, A., 1943. On a measure of divergence between two statistical populations defined by their probability distribution. Bull. Calcutta Math. Soc.

Bianconi, F., Fernández, A., 2014. An appendix to "Texture databases–A comprehensive survey". Pattern Recognition Letters 45, 33–38.

Bouremoum, S., Protonotarios, E. D., Griffin, L. D., 2015. Quantifying texture scale. Acta Stereologica.

Boutellaa, E., López, M. B., Ait-Aoudia, S., Feng, X., Hadid, A., Aug. 2017. Kinship Verification from Videos using Spatio-Temporal Texture Features and Deep Learning. arXiv:1708.04069 [cs]ArXiv: 1708.04069.

Bradley, R. A., Terry, M. E., 1952. Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. Biometrika 39 (3/4), 324–345.

Brodatz, P., 1966a. Textures. Dover, New York.

Brodatz, P., 1966b. Textures: a photographic album for artists and designers. Vol. 66. Dover New York.

Brox, T., Weickert, J., 2004. A TV flow based local scale measure for texture discrimination. In: Computer Vision-ECCV 2004. Springer, pp. 578–590.

Bruynooghe, M., Jun. 2006. Mammographic Mass Detection Using Unsupervised Clustering in Synergy with a Parcimonious Supervised Rule-Based Classifier. In:

Astley, S. M., Brady, M., Rose, C., Zwiggelaar, R. (Eds.), Digital Mammography. No. 4046 in Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 68–75.

Campbell, F. W., Robson, J. G., Aug. 1968. Application of fourier analysis to the visibility of gratings. The Journal of Physiology 197 (3), 551–566.

Castellano, G., Bonilha, L., Li, L. M., Cendes, F., Dec. 2004. Texture analysis of medical images. Clinical Radiology 59 (12), 1061–1069.

Chan, A. B., Vasconcelos, N., 2007. Classifying Video with Kernel Dynamic Textures. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–6.

Chang, T., Kuo, C. J., 1993. Texture analysis and classification with tree-structured wavelet transform. Image Processing, IEEE Transactions on 2 (4), 429–441.

Chellappa, R., Chatterjee, S., 1985. Classification of textures using Gaussian Markov random fields. Acoustics, Speech and Signal Processing, IEEE Transactions on 33 (4), 959–963.

Chen, D.-R., Chang, R.-F., Kuo, W.-J., Chen, M.-C., Huang, Y.-L., 2002. Diagnosis of breast tumors with sonographic texture analysis using wavelet transform and neural networks. Ultrasound in medicine & biology 28 (10), 1301–1310.

Cimpoi, M., Maji, S., Kokkinos, I., Vedaldi, A., May 2016. Deep Filter Banks for Texture Recognition, Description, and Segmentation. International Journal of Computer Vision 118 (1), 65–94.

Cimpoi, M., Maji, S., Vedaldi, A., 2015. Deep Filter Banks for Texture Recognition and Segmentation. pp. 3828–3836.

Ciocca, G., Corchs, S., Gasparini, F., 2015a. Complexity Perception of Texture Images. In: International Conference on Image Analysis and Processing. Springer, pp. 119–126.

Ciocca, G., Corchs, S., Gasparini, F., Bricolo, E., Tebano, R., 2015b. Does color influence image complexity perception? In: International Workshop on Computational Color Imaging. Springer, pp. 139–148.

Clarke, A. D., Halley, F., Newell, A. J., Griffin, L. D., Chantler, M. J., 2011. Perceptual Similarity: A Texture Challenge. In: BMVC. Citeseer, pp. 1–10.

Clausi, D. A., Jan. 2002. An analysis of co-occurrence texture statistics as a function of grey level quantization. Canadian Journal of Remote Sensing 28 (1), 45–62.

Coggins, J. M., 1983. A framework for texture analysis based on spatial filtering.

Cohen, J., 1968. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. Psychological bulletin 70 (4), 213.

Cohen, J., others, 1960. A coefficient of agreement for nominal scales. Educational and psychological measurement 20 (1), 37–46.

Conners, R. W., Harlow, C. A., 1980. Toward a structural textural analyzer based on statistical methods. Computer Graphics and Image Processing 12 (3), 224–256.

Corrigall, D., Walther, R. F., Rodriguez, L., Fichelson, P., Pichaud, F., 2007. Hedgehog signaling is a principal inducer of Myosin-II-driven cell ingression in Drosophila epithelia. Developmental cell 13 (5), 730–742.

Costantini, R., Sbaiz, L., Susstrunk, S., 2008. Higher Order SVD Analysis for Dynamic Texture Synthesis. IEEE Transactions on Image Processing 17 (1), 42–52.

Crosier, M., Griffin, L. D., 2010. Using basic image features for texture classification. International Journal of Computer Vision 88 (3), 447–460.

Cross, G. R., Jain, A. K., 1983. Markov random field texture models. Pattern Analysis and Machine Intelligence, IEEE Transactions on (1), 25–39.

Cula, O. G., Dana, K. J., Aug. 2004. 3d texture recognition using bidirectional feature histograms. International Journal of Computer Vision 59 (1), 33–60, wOS:000221100600002.

Cula, O. G., Dana, K. J., Murphy, F. P., Rao, B. K., Dec. 2004. Bidirectional imaging and modeling of skin texture. Ieee Transactions on Biomedical Engineering 51 (12), 2148–2159, wOS:000225285600008.

Dana, K. J., Van Ginneken, B., Nayar, S. K., Koenderink, J. J., 1999. Reflectance and texture of real-world surfaces. ACM Transactions on Graphics (TOG) 18 (1), 1–34.

David, H. A., 1963. The method of paired comparisons. Vol. 12. London.

de Siqueira, F. R., Schwartz, W. R., Pedrini, H., 2013. Multi-scale gray level co-occurrence matrices for texture description. Neurocomputing 120, 336–345.

Derpanis, K. G., Wildes, R. P., Jun. 2010. Dynamic texture recognition based on distributions of spacetime oriented structure. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 191–198.

Do, M. N., Vetterli, M., 2002. Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. Image Processing, IEEE Transactions on 11 (2), 146–158.

Dong, J., Chantler, M., 2004. On the relations between three methods for representing 3d surface textures under arbitrary illumination directions. In: Computer and Information Technology, 2004. CIT'04. The Fourth International Conference on. IEEE, pp. 807–812.

Dong, J., Ma, S., Li, L., Yu, Z., 2007. Hole filling on three-dimensional surface texture. In: 2007 IEEE International Conference on Multimedia and Expo. IEEE, pp. 1299–1302.

Doretto, G., Chiuso, A., Wu, Y. N., Soatto, S., Feb. 2003a. Dynamic textures. International Journal of Computer Vision 51 (2), 91–109, wOS:000179809200001.

Doretto, G., Cremers, D., Favaro, P., Soatto, S., 2003b. Dynamic texture segmentation. In: Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on. IEEE, pp. 1236–1242.

Doretto, G., Soatto, S., 2003. Editable dynamic textures. In: 2003 Ieee Computer Society Conference on Computer Vision and Pattern Recognition, Vol Ii, Proceedings. pp. 137–142, wOS:000184081500019.

Drbohlav, O., Chantler, M., 2005. Illumination-invariant texture classification using single training images. In: Texture. pp. 31–36.

Dunn, D., Higgins, W. E., 1995. Optimal Gabor filters for texture segmentation. Image Processing, IEEE Transactions on 4 (7), 947–964.

Efros, A. A., Freeman, W. T., 2001. Image Quilting for Texture Synthesis and Transfer. In: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques. SIGGRAPH '01. ACM, New York, NY, USA, pp. 341–346.

Efros, A. A., Leung, T. K., 1999. Texture synthesis by non-parametric sampling. In: Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on. Vol. 2. IEEE, pp. 1033–1038.

Ehrenstein, W. H., Ehrenstein, A., 1999. Psychophysical methods. In: Modern techniques in neuroscience research. Springer, pp. 1211–1241.

Esgiar, A. N., Naguib, R. N. G., Sharif, B. S., Bennett, M. K., Murray, A., Mar. 2002. Fractal analysis in the detection of colonic cancer images. IEEE Transactions on Information Technology in Biomedicine 6 (1), 54–58.

Fazekas, S., Chetverikov, D., Aug. 2007. Analysis and performance evaluation of optical flow features for dynamic texture recognition. Signal Processing: Image Communication 22 (7–8), 680–691.

Fechner, G., 1966. Elements of psychophysics. Vol. I.

Filip, J., Haindl, M., Chetverikov, D., 2006. Fast Synthesis of Dynamic Colour Textures. In: 18th International Conference on Pattern Recognition (ICPR'06). Vol. 4. pp. 25–28.

Forsyth, D., 2002. Shape from texture without boundaries. Computer Vision—ECCV 2002, 43–66.

Fritz, M., Hayman, E., Caputo, B., Eklundh, J.-O., 2004. The kth-tips database.

Georgeson, M. A., 1979. Spatial Fourier analysis and human vision. In: Tutorial essays in psychology. Vol. 2. Lawrence Erlbaum Hilsdale, NJ, pp. 39–88.

Gescheider, G., 1997. Chapter 3. The Classical Psychophysical Methods. Psychophysics: the fundamentals. 3rd ed. Mahwah: Lawrence Erlbaum Associates.

Gescheider, G. A., 1985. Psychophysics: method. Theory and Application (New York: Laurence Erlbaum).

Gescheider, G. A., 1988. Psychophysical Scaling. Annual Review of Psychology 39 (1), 169–200.

Geusebroek, J. M., Burghouts, G. J., 2009. Amsterdam Library of Textures (ALOT).

Ghanem, B., Ahuja, N., Sep. 2010. Maximum Margin Distance Learning for Dynamic Texture Recognition. In: Daniilidis, K., Maragos, P., Paragios, N. (Eds.), Computer Vision – ECCV 2010. No. 6312 in Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 223–236.

Gotlieb, C. C., Kreyszig, H. E., 1990. Texture descriptors based on co-occurrence matrices. Computer Vision, Graphics, and Image Processing 51 (1), 70–86.

Gower, J. C., Dijksterhuis, G. B., 2004. Procrustes problems. Vol. 3. Oxford University Press Oxford.

Greenwood, S., Struhl, G., 1999. Progression of the morphogenetic furrow in the Drosophila eye: the roles of Hedgehog, Decapentaplegic and the Raf pathway. Development 126 (24), 5795–5808.

Griffin, L. D., 2007. The second order local-image-structure solid. Pattern Analysis and Machine Intelligence, IEEE Transactions on 29 (8), 1355–1366.

Griffin, L. D., 2008. Symmetries of 1-D images. Journal of Mathematical Imaging and Vision 31 (2-3), 157–164.

Griffin, L. D., 2009. Symmetries of 2-D Images: cases without periodic translations. Journal of Mathematical Imaging and Vision 34 (3), 259–269.

Griffin, L. D., Lillholm, M., Crosier, M., van Sande, J., 2009. Basic image features (BIFs) arising from approximate symmetry type. In: Scale Space and Variational Methods in Computer Vision. Springer, pp. 343–355.

Grigorescu, S., Petkov, N., Sep. 2003. Texture analysis using Renyi's generalized entropies. In: 2003 International Conference on Image Processing, 2003. ICIP 2003. Proceedings. Vol. 1. pp. I–241–4 vol.1.

Grigorescu, S. E., Petkov, N., Kruizinga, P., 2002. Comparison of texture features based on Gabor filters. Image Processing, IEEE Transactions on 11 (10), 1160–1167.

Guo, Y., Zhao, G., Zhou, Z., Pietikäinen, M., 2013. Video texture synthesis with multi-frame LBP-TOP and diffeomorphic growth model. IEEE Transactions on Image Processing 22 (10), 3879–3891.

Guo, Z., Zhang, L., Zhang, D., 2010. A completed modeling of local binary pattern operator for texture classification. Image Processing, IEEE Transactions on 19 (6), 1657–1663.

Hand, D., Dec. 2004. Measurement Theory and Practice. {A Hodder Arnold Publication}.

Hand, D. J., 1996. Statistics and the Theory of Measurement. Journal of the Royal Statistical Society. Series A (Statistics in Society) 159 (3), 445–492.

Haralick, R. M., 1979. Statistical and structural approaches to texture. Proceedings of the IEEE 67 (5), 786–804.

Harris, C., Stephens, M., 1988. A combined corner and edge detector. In: Alvey vision conference. Vol. 15. Citeseer, p. 50.

Hawkins, J. K., 1970. Textural properties for pattern recognition. Picture processing and psychopictorics, 347–370.

Hayman, E., Caputo, B., Fritz, M., Eklundh, J. O., 2004. On the significance of real-world conditions for material classification. In: Pajdla, T., Matas, J. (Eds.), Computer Vision - Eccv 2004, Pt 4. Vol. 2034. pp. 253–266.

Hays, J., Leordeanu, M., Efros, A. A., Liu, Y., 2006. Discovering texture regularity as a higher-order correspondence problem. In: Computer Vision–ECCV 2006. Springer, pp. 522–535.

Hong, B.-W., Soatto, S., Ni, K., Chan, T., 2008. The scale of a texture and its application to segmentation. In: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, pp. 1–8.

Hossain, S., Serikawa, S., Nov. 2013. Texture databases – A comprehensive survey. Pattern Recognition Letters 34 (15), 2007–2022.

Hsiao, J. Y., Sawchuk, A. A., 1989. Unsupervised textured image segmentation using feature smoothing and probabilistic relaxation techniques. Computer Vision, Graphics, and Image Processing 48 (1), 1–21.

Idrissa, M., Acheroy, M., 2002. Texture classification using Gabor filters. Pattern Recognition Letters 23 (9), 1095–1102.

Jaccard, N., Szita, N., Griffin, L. D., Sep. 2017. Segmentation of phase contrast microscopy images based on multi-scale local Basic Image Features histograms. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization 5 (5), 359–367.

Jain, A. K., Farrokhnia, F., 1991. Unsupervised texture segmentation using Gabor filters. Pattern recognition 24 (12), 1167–1186.

Jan, S.-R., Hsueh, Y.-C., 1998. Window-size determination for granulometrical structural texture classification. Pattern Recognition Letters 19 (5), 439–446.

Jayasakthi, J., Rajaselvi, V. M., 2016. Detection of Macular Degeneration in Retinal Images Based on Texture Segmentation. In: Suresh, L. P., Panigrahi, B. K. (Eds.), Proceedings of the International Conference on Soft Computing Systems. No. 397 in Advances in Intelligent Systems and Computing. Springer India, pp. 439–449.

Ji, H., Yang, X., Ling, H., Xu, Y., 2013. Wavelet domain multifractal analysis for static and dynamic texture classification. IEEE Transactions on Image Processing 22 (1), 286–299.

Jian, M., Dong, J., 2011. Capture and fusion of 3d surface texture. Multimedia Tools and Applications 53 (1), 237–251.

Jojic, N., Frey, B., Kannan, A., Oct. 2003. Epitomic analysis of appearance and shape. In: Ninth IEEE International Conference on Computer Vision, 2003. Proceedings. pp. 34–41 vol.1.

Kadir, T., Brady, M., 2001. Saliency, scale and image description. International Journal of Computer Vision 45 (2), 83–105.

Ke, Y., Sukthankar, R., 2004. PCA-SIFT: A more distinctive representation for local image descriptors. In: Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on. Vol. 2. IEEE, pp. II–506.

Keller, J. M., Chen, S., Crownover, R. M., 1989. Texture description and segmentation through fractal geometry. Computer Vision, Graphics, and Image Processing 45 (2), 150–166.

Kim, K. I., Jung, K., Park, S. H., Kim, H. J., 2002. Support vector machines for texture classification. IEEE transactions on pattern analysis and machine intelligence 24 (11), 1542–1550.

Koenderink, J. J., 1984. The structure of images. Biological cybernetics 50 (5), 363–370.

Kong, H., Gurcan, M., Belkacem-Boussaid, K., 2011. Partitioning histopathological images: an integrated framework for supervised color-texture segmentation and cell splitting. IEEE transactions on medical imaging 30 (9), 1661–1677.

Kong, S., Wang, D., 2012. Multi-level feature descriptor for robust texture classification via locality-constrained collaborative strategy. arXiv preprint arXiv:1203.0488.

Koss, J. E., Newman, F. D., Johnson, T. K., Kirch, D. L., 1999. Abdominal organ segmentation using texture transforms and a hopfield neural network. IEEE Transactions on medical imaging 18 (7), 640–648.

Kumar, P. V., Kumar, P. S., Madhuri, N., Devi, M. U., 2016. Stone Image Classification Based on Overlapped 5-bit T-Patterns occurrence on 5-by-5 Sub Images. International Journal of Electrical and Computer Engineering (IJECE) 6 (3).

Lazebnik, S., Schmid, C., Ponce, J., 2005a. A maximum entropy framework for part-based texture and object recognition. In: Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1. Vol. 1. IEEE, pp. 832–838.

Lazebnik, S., Schmid, C., Ponce, J., 2005b. A sparse texture representation using local affine regions. Pattern Analysis and Machine Intelligence, IEEE Transactions on 27 (8), 1265–1278.

LeCun, Y., Bengio, Y., Hinton, G., May 2015. Deep learning. Nature 521 (7553), 436–444.

Lee, J., Zee, B. C. Y., Li, Q., 2013. Detection of neovascularization based on fractal and texture analysis with interaction effects in diabetic retinopathy. PloS one 8 (12), e75699.

Leu, J.-G., 2001. On indexing the periodicity of image textures. Image and Vision Computing 19 (13), 987–1000.

Leung, T., Malik, J., 1996. Detecting, localizing and grouping repeated scene elements from an image. Computer Vision—ECCV'96, 546–555.

Leung, T., Malik, J., 2001. Representing and recognizing the visual appearance of materials using three-dimensional textons. International Journal of Computer Vision 43 (1), 29–44.

Li, S. Z., 2012. Markov random field modeling in computer vision. Springer Science & Business Media.

Li, Y., Wang, T., Shum, H.-Y., 2002. Motion Texture: A Two-level Statistical Model for Character Motion Synthesis. In: Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques. SIGGRAPH '02. ACM, New York, NY, USA, pp. 465–472.

Liao, S., Law, M. W., Chung, A., 2009. Dominant local binary patterns for texture classification. Image Processing, IEEE Transactions on 18 (5), 1107–1118.

Lin, H.-C., Wang, L.-L., Yang, S.-N., 1997. Extracting periodicity of a regular texture based on autocorrelation functions. Pattern Recognition Letters 18 (5), 433–443.

Lin, W.-C., Hays, J., Wu, C., Liu, Y., Kwatra, V., 2006a. Quantitative evaluation of near regular texture synthesis algorithms. In: Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. Vol. 1. IEEE, pp. 427–434.

Lin, W.-C., Liu, Y., 2007. A lattice-based MRF model for dynamic near-regular texture tracking. Pattern Analysis and Machine Intelligence, IEEE Transactions on 29 (5), 777–792.

Lin, Y.-C., Tsai, Y.-P., Hung, Y.-P., Shih, Z.-C., 2006b. Comparison between immersion-based and toboggan-based watershed image segmentation. IEEE Transactions on image processing 15 (3), 632–640.

Lindeberg, T., 1994. Scale selection for differential operators. Springer.

Liu, L., Fieguth, P., 2012. Texture classification from random features. IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (3), 574–586.

Liu, L., Fieguth, P., Kuang, G., Zha, H., Nov. 2011. Sorted Random Projections for robust texture classification. In: 2011 IEEE International Conference on Computer Vision (ICCV). pp. 391–398.

Liu, L., Zhao, L., Long, Y., Kuang, G., Fieguth, P., 2012. Extended local binary patterns for texture classification. Image and Vision Computing 30 (2), 86–99.

Liu, Y., Collins, R. T., Tsin, Y., 2004a. A computational model for periodic pattern perception based on frieze and wallpaper groups. Pattern Analysis and Machine Intelligence, IEEE Transactions on 26 (3), 354–371.

Liu, Y. X., Lin, W. C., Hays, J., Aug. 2004b. Near-regular texture analysis and manipulation. Acm Transactions on Graphics 23 (3), 368–376.

Lizarraga-Morales, R. A., Sanchez-Yanez, R. E., Ayala-Ramirez, V., 2011. Homogeneity cues for texel size estimation of periodic and near-periodic textures. In: Pattern Recognition. Springer, pp. 220–229.

Lobay, A., Forsyth, D. A., 2004. Recovering shape and irradiance maps from rich dense texton fields. In: Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on. Vol. 1. IEEE, pp. I–I.

Lorigo, L. M., Faugeras, O., Grimson, W. E. L., Keriven, R., Kikinis, R., 1998. Segmentation of bone in clinical knee MRI using texture-based geodesic active contours. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 1195–1204.

Lowe, D. G., 1999. Object recognition from local scale-invariant features. In: Computer vision, 1999. The proceedings of the seventh IEEE international conference on. Vol. 2. Ieee, pp. 1150–1157.

Lowe, D. G., Nov. 2004. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 60 (2), 91–110.

Luce, D. R., Krumhansl, C. L., 1988. Measurement, scaling and psychophysics. In: Stevens' handbook of experimental psychology. Vol. 1. pp. 3–74.

Malik, J., Belongie, S., Leung, T., Shi, J., 2001. Contour and texture analysis for image segmentation. International journal of computer vision 43 (1), 7–27.

Mallikarjuna, P., Targhi, A. T., 2006. KTH-TIPS2 Image Database.

Mandelbrot, B. B., 1983. The fractal geometry of nature. Vol. 173. Macmillan.

Manjunath, B. S., Ma, W.-Y., 1996. Texture features for browsing and retrieval of image data. Pattern Analysis and Machine Intelligence, IEEE Transactions on 18 (8), 837–842.

Materka, A., Strzelecki, M., others, 1998. Texture analysis methods–a review. Technical university of lodz, institute of electronics, COST B11 report, Brussels, 9–11.

McKenna, F. P., Feb. 1985. Another look at the 'new psychophysics'. British Journal of Psychology 76 (1), 97–109.

Mellor, M., Hong, B.-W., Brady, M., 2008. Locally rotation, contrast, and scale invariant descriptors for texture analysis. Pattern Analysis and Machine Intelligence, IEEE Transactions on 30 (1), 52–61.

Mäenpää, T., Pietikäinen, M., 2004. Classification with color and texture: jointly or separately? Pattern recognition 37 (8), 1629–1640.

Mosteller, F., 2006. Remarks on the Method of Paired Comparisons: I. The Least Squares Solution Assuming Equal Standard Deviations and Equal Correlations. In: Selected Papers of Frederick Mosteller. Springer Series in Statistics. Springer, New York, NY, pp. 157–162.

Murni Adnin, M., Smith, W., 2001. Mayang's Free Texture Library.

Mylonas, D., MacDonald, L., Wuerger, S., 2010. Towards an online color naming model. In: Color and Imaging Conference. Vol. 2010. Society for Imaging Science and Technology, pp. 140–144.

Nadler, M., Smith, E. P., 1993. Pattern recognition engineering. Wiley-interscience.

Nowak, E., Jurie, F., Triggs, B., 2006. Sampling strategies for bag-of-features image classification. In: Computer Vision–ECCV 2006. Springer, pp. 490–503.

Oh, G., Lee, S., Shin, S. Y., 1999. Fast determination of textural periodicity using distance matching function. Pattern Recognition Letters 20 (2), 191–197.

Ojala, T., Maenpaa, T., Pietikainen, M., Viertola, J., Kyllonen, J., Huovinen, S., 2002a. Outex - New framework for empirical evaluation of texture analysis algorithms. In: Kasturi, R., Laurendeau, D., Suen, C. (Eds.), 16th International Conference on Pattern Recognition, Vol I, Proceedings. pp. 701–706, wOS:000177847900169.

Ojala, T., Pietikainen, M., Maenpaa, T., 2002b. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. Pattern Analysis and Machine Intelligence, IEEE Transactions on 24 (7), 971–987.

Ojala, T., Pietikäinen, M., 1999. Unsupervised texture segmentation using feature distributions. Pattern Recognition 32 (3), 477–486.

Ojala, T., Pietikäinen, M., Harwood, D., 1996. A comparative study of texture measures with classification based on featured distributions. Pattern recognition 29 (1), 51–59.

Ojansivu, V., Heikkilä, J., 2008. Blur insensitive texture classification using local phase quantization. In: International conference on image and signal processing. Springer, pp. 236–243.

Ouyang, W., Wang, X., 2013. Joint Deep Learning for Pedestrian Detection. pp. 2056–2063.

Paget, R., Longstaff, D., 1995. Texture synthesis via a non-parametric Markov random Field. Proceedings of DICTA-95, Digital Image Computing: Techniques and Applications 1, 547–552.

Panjwani, D. K., Healey, G., 1995. Markov random field models for unsupervised segmentation of textured color images. Pattern Analysis and Machine Intelligence, IEEE Transactions on 17 (10), 939–954.

Pareek, N. K., Patidar, V., Sud, K. K., 2006. Image encryption using chaotic logistic map. Image and Vision Computing 24 (9), 926–934.

Park, M., 2010. Efficient mean-shift belief propagation and its applications in computer vision. Pennsylvania State University.

Park, M., Brocklehurst, K., Collins, R. T., Liu, Y., 2009. Deformed lattice detection in real-world images using mean-shift belief propagation. Pattern Analysis and Machine Intelligence, IEEE Transactions on 31 (10), 1804–1816.

Parkkinen, J., Selkäinaho, K., Oja, E., 1990. Detecting texture periodicity from the cooccurrence matrix. Pattern Recognition Letters 11 (1), 43–50.

Peckinpaugh, S. H., 1991. An improved method for computing gray-level cooccurrence matrix based texture measures. CVGIP: Graphical Models and Image Processing 53 (6), 574–580.

Pentland, A. P., 1984. Fractal-based description of natural scenes. Pattern Analysis and Machine Intelligence, IEEE Transactions on (6), 661–674.

Permuter, H., Francos, J., Jermyn, I. H., 2003. Gaussian mixture models of texture and colour for image database retrieval. In: Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on. Vol. 3. IEEE, pp. III–569.

Peteri, R., Chetverikov, D., 2005. Dynamic texture recognition using normal flow and texture regularity. In: Marques, J. S., PerezdelaBlanca, N., Pina, P. (Eds.),

Pattern Recognition and Image Analysis, Pt 2, Proceedings. Vol. 3523. pp. 223–230, wOS:000230027000028.

Peteri, R., Fazekas, S., Huiskes, M. J., Sep. 2010. DynTex: A comprehensive database of dynamic textures. Pattern Recognition Letters 31 (12), 1627–1632, wOS:000281368200017.

Picard, R., Graczyk, C., Mann, S., Wachman, J., Picard, L., Campbell, L., Jan. 2010. VisTex vision texture database. ResearchGate.

Pichaud, F., 2014. Transcriptional regulation of tissue organization and cell morphogenesis: The fly retina as a case study. Developmental biology 385 (2), 168–178.

Päivärinta, J., Rahtu, E., Heikkilä, J., May 2011. Volume Local Phase Quantization for Blur-Insensitive Dynamic Texture Classification. In: Heyden, A., Kahl, F. (Eds.), Image Analysis. No. 6688 in Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 360–369.

Po, D. D., Do, M. N., 2003. Directional multiscale statistical modeling of images. In: Optical Science and Technology, SPIE's 48th Annual Meeting. International Society for Optics and Photonics, pp. 69–79.

Portilla, J., Simoncelli, E. P., 2000. A parametric texture model based on joint statistics of complex wavelet coefficients. International Journal of Computer Vision 40 (1), 49–70.

Protonotarios, E. D., Baum, B., Johnston, A., Hunter, G. L., Griffin, L. D., 2014. An absolute interval scale of order for point patterns. Journal of The Royal Society Interface 11 (99), 20140342.

Quan, Y., Xu, Y., Sun, Y., Luo, Y., 2014. Lacunarity analysis on image patterns for texture classification. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 160–167.

Ramakrishnan, S., Selvan, S., 2008. Classification of soil texture based on wavelet domain singular values. International Journal of Tomography & Statistics 9, 33–51.

Rao, A. R., Lohse, G. L., 1993. Identifying high level features of texture perception. CVGIP: Graphical Models and Image Processing 55 (3), 218–233.

Ren, J., Jiang, X., Yuan, J., 2013. Dynamic texture recognition using enhanced LBP features. In: ICASSP. pp. 2400–2404.

Robertson, F., Pinal, N., Fichelson, P., Pichaud, F., 2012. Atonal and EGFR signalling orchestrate rok-and Drak-dependent adherens junction remodelling during ommatidia morphogenesis. Development 139 (18), 3432–3441.

Rogers, T. W., Jaccard, N., Morton, E. J., Griffin, L. D., Jan. 2017. Automated X-ray image analysis for cargo security: Critical review and future promise. Journal of X-Ray Science and Technology 25 (1), 33–56.

Roland, K., Meerwald, P., 2009. Salzburg Texture Image Database (STex).

Rosenfeld, A., Kak, A. C., Kak, A. C., Aug. 1982. Digital Picture Processing, Volume 1, Second Edition, 2nd Edition. Morgan Kaufmann, New York.

Saisan, P., Doretto, G., Wu, Y. N., Soatto, S., 2001. Dynamic texture recognition. In: Jacobs, A., Baldwin, T. (Eds.), 2001 Ieee Computer Society Conference on Computer Vision and Pattern Recognition, Vol 2, Proceedings. pp. 58–63.

Sarkar, N., Chaudhuri, B. B., 1992. An efficient approach to estimate fractal dimension of textural images. Pattern Recognition 25 (9), 1035–1041.

Schaffalitzky, F., Zisserman, A., 1999. Geometric grouping of repeated elements within images. Shape, contour and grouping in computer vision, 81–81.

Schattschneider, D., 1978. The plane symmetry groups: their recognition and notation. The American Mathematical Monthly 85 (6), 439–450.

Schindler, G., Krishnamurthy, P., Lublinerman, R., Liu, Y., Dellaert, F., 2008. Detecting and matching repeated patterns for automatic geo-tagging in urban environments. In: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, pp. 1–7.

Sebastian, T. B., Tek, H., Crisco, J. J., Kimia, B. B., 2003. Segmentation of carpal bones from CT images using skeletally coupled deformable models. Medical Image Analysis 7 (1), 21–45.

Setti, F., Conigliaro, D., Tobanelli, M., Cristani, M., 2017. Count on me: learning to count on a single image. IEEE Transactions on Circuits and Systems for Video Technology.

Sharma, N., Ray, A. K., Sharma, S., Shukla, K. K., Pradhan, S., Aggarwal, L. M., others, 2008. Segmentation and classification of medical images using texture-primitive features: Application of BAM-type artificial neural network. Journal of medical physics 33 (3), 119.

Shen, D., Wu, G., Suk, H.-I., 2017. Deep Learning in Medical Image Analysis. Annual Review of Biomedical Engineering 19 (1), 221–248.

Sifre, L., Mallat, S., 2013. Rotation, scaling and deformation invariant scattering for texture discrimination. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1233–1240.

Sklansky, J., 1978. Image segmentation and feature extraction. Systems, Man and Cybernetics, IEEE Transactions on 8 (4), 237–247.

Smith, G., Burns, I., 1997. Meastex image texture database and test suite. URL: http://www. cssip. uq. edu. au/meastex/meastex. html,[Last accessed on July 31, 2005].

Smith, J. R., Chang, S.-F., 1994. Transform features for texture classification and discrimination in large image databases. In: Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference. Vol. 3. IEEE, pp. 407–411.

Soatto, S., Doretto, G., Wu, Y. N., 2001. Dynamic textures. In: Eighth IEEE International Conference on Computer Vision, 2001. ICCV 2001. Proceedings. Vol. 2. pp. 439–446 vol.2.

Sorensen, L., Shaker, S. B., De Bruijne, M., 2010. Quantitative analysis of pulmonary emphysema using local binary patterns. IEEE transactions on medical imaging 29 (2), 559–569.

Sotoca, J. M., Iñesta, J. M., Belmonte, M. A., 2003. Hand bone segmentation in radioabsorptiometry images for computerised bone mass assessment. Computerized Medical Imaging and Graphics 27 (6), 459–467.

Stam, J., 1997. Aperiodic texture mapping. European Research Consortium for Informatics and Mathematics.

Starovoitov, V. V., Jeong, S.-Y., Park, R.-H., 1998. Texture periodicity detection: Features, properties, and comparisons. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on 28 (6), 839–849.

Stevens, S., 1946. On the theory of scales of measurement. Science 103 (2684), 677–680.

Stevens, S., 1951. Mathematics, measurement and psychophysics. Handbook of Experimental Pscyhology, Wiley, New York 1, 1–49.

Stevens, S. S., 1953. On the brightness of lights and loudness of sounds. Science 118, 576.

Stevens, S. S., 1955. The measurement of loudness. The Journal of the Acoustical Society of America 27 (5), 815–829.

Stevens, S. S., 1975. Psychophysics: Introduction to its perceptual, neural and social prospects. Routledge.

Sun, Y., Wang, X., Tang, X., 2014. Deep Learning Face Representation from Predicting 10,000 Classes. pp. 1891–1898.

Swanson, W. H., Birch, E. E., Sep. 1992. Extracting thresholds from noisy psychophysical data. Perception & Psychophysics 51 (5), 409–422.

Tamura, H., Mori, S., Yamawaki, T., 1978. Textural features corresponding to visual perception. Systems, Man and Cybernetics, IEEE Transactions on 8 (6), 460–473.

Tanay, T., Griffin, L., Aug. 2016. A Boundary Tilting Persepective on the Phenomenon of Adversarial Examples. arXiv:1608.07690 [cs, stat]ArXiv: 1608.07690.

Targhi, A. T., Geusebroek, J.-M., Zisserman, A., 2008. Texture classification with minimal training images. In: Pattern Recognition, 2008. ICPR 2008. 19th International Conference on. IEEE, pp. 1–4.

Teuner, A., Pichler, O., Hosticka, B. J., 1995. Unsupervised texture segmentation of images using tuned matched Gabor filters. Image Processing, IEEE Transactions on 4 (6), 863–870.

Thomas, H., 1981. Estimation in the power law. Psychometrika 46 (1), 29–34.

Thurstone, L. L., 1927. A law of comparative judgment. Psychological review 34 (4), 273.

Tiwari, D., Tyagi, V., 2017. Improved Weber's law based local binary pattern for dynamic texture recognition. Multimedia Tools and Applications 76 (5), 6623–6640.

Todorovic, S., Ahuja, N., 2009. Texel-based texture segmentation. In: Computer Vision, 2009 IEEE 12th International Conference on. IEEE, pp. 841–848.

Torgerson, W. S., 1958. Theory and methods of scaling.

Torii, A., Sivic, J., Pajdla, T., Okutomi, M., 2013. Visual place recognition with repetitive structures. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 883–890.

Towsend, J. T., Ashby, F. G., 1984. Measurement scales and statistics: The misconception misconceived. Psychological Bulletin 96(2), 394–401.

Tsukida, K., Gupta, M. R., May 2011. How to Analyze Paired Comparison Data. Tech. Rep. UWEETR-2011-0004, WASHINGTON UNIV SEATTLE DEPT OF ELECTRICAL ENGINEERING, WASHINGTON UNIV SEATTLE DEPT OF ELECTRICAL ENGINEERING.

Tuceryan, M., Jain, A. K., 1998. Texture analysis. The handbook of pattern recognition and computer vision 2, 207–248.

Turina, A., Tuytelaars, T., Van Gool, L., 2001. Efficient grouping under perspective skew. In: Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. Vol. 1. IEEE, pp. I–I.

Ulaby, F. T., Kouyate, F., Brisco, B., Williams, T. L., 1986. Textural infornation in SAR images. Geoscience and Remote Sensing, IEEE Transactions on (2), 235–245.

Unser, M., 1995. Texture classification and segmentation using wavelet frames. Image Processing, IEEE Transactions on 4 (11), 1549–1560.

Various, A., 1999. PhoTex Database.

Various, A., 2006. UMD HR Texture Dataset.

Varma, M., Garg, R., 2007. Locally invariant fractal features for statistical texture classification. In: Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on. IEEE, pp. 1–8.

Varma, M., Zisserman, A., 2003. Texture classification: Are filter banks necessary? In: Computer vision and pattern recognition, 2003. Proceedings. 2003 IEEE computer society conference on. Vol. 2. IEEE, pp. II–691.

Voronoï, G., 1908. Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième mémoire. Recherches sur les parallélloèdres primitifs. Journal für die reine und angewandte Mathematik 134, 198–287.

Wang, L., He, D.-C., 1990. Texture classification using texture spectrum. Pattern Recognition 23 (8), 905–910.

Wang, Y., Gao, X., Li, J., 2007. A feature analysis approach to mass detection in mammography based on RF-SVM. In: 2007 IEEE International Conference on Image Processing. Vol. 5. IEEE, pp. V–9.

Wassenberg, J., Middelmann, W., Sanders, P., 2009. An efficient parallel algorithm for graph-based image segmentation. In: International Conference on Computer Analysis of Images and Patterns. Springer, pp. 1003–1010.

Weber, A. G., 1997. USC SIPI Image Database.

Wei, L.-Y., Levoy, M., 2000. Fast texture synthesis using tree-structured vector quantization. In: Proceedings of the 27th annual conference on Computer graphics and interactive techniques. ACM Press/Addison-Wesley Publishing Co., pp. 479–488.

Weldon, T. P., Higgins, W. E., Dunn, D. F., 1996. Efficient Gabor filter design for texture segmentation. Pattern Recognition 29 (12), 2005–2015.

Wichmann, F. A., Hill, N. J., 2001. The psychometric function: I. Fitting, sampling, and goodness of fit. Perception & psychophysics 63 (8), 1293–1313.

Won, C. S., Derin, H., 1992. Unsupervised segmentation of noisy and textured images using Markov random fields. CVGIP: Graphical Models and Image Processing 54 (4), 308–328.

Wu, Y., Yang, G., Jin, H., Noonan, J. P., 2012. Image encryption using the two-dimensional logistic chaotic map. Journal of Electronic Imaging 21 (1), 013014–1.

Xian, G.-m., 2010. An identification method of malignant and benign liver tumors from ultrasonography based on GLCM texture features and fuzzy SVM. Expert Systems with Applications 37 (10), 6737–6741.

Xu, Y., Huang, S., Ji, H., Fermüller, C., 2012. Scale-space texture description on SIFT-like textons. Computer Vision and Image Understanding 116 (9), 999–1013.

Xu, Y., Yang, X., Ling, H., Ji, H., 2010. A new texture descriptor using multifractal analysis in multi-orientation wavelet pyramid. In: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, pp. 161–168.

Ylioinas, J., Hadid, A., Guo, Y., Pietikäinen, M., 2013. Efficient image appearance description using dense sampling based local binary patterns. In: Computer Vision–ACCV 2012. Springer, pp. 375–388.

Yuan, L., Wen, F., Liu, C., Shum, H. Y., 2004. Synthesizing dynamic texture with closed-loop linear dynamic system. In: Pajdla, T., Matas, J. (Eds.), Computer Vision - Eccv 2004, Pt 2. Vol. 3022. pp. 603–616, wOS:000221519200048.

Zhang, G., Wang, W., Moon, J., Pack, J. K., Jeon, S. I., 2011. A Review of Breast Tissue Classification in Mammograms. In: Proceedings of the 2011 ACM Symposium on Research in Applied Computation. RACS '11. ACM, New York, NY, USA, pp. 232–237.

Zhang, H., Berg, A. C., Maire, M., Malik, J., 2006. SVM-KNN: Discriminative nearest neighbor classification for visual category recognition. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). Vol. 2. IEEE, pp. 2126–2136.

Zhang, J., Liang, J., Zhang, C., Zhao, H., Jan. 2015. Scale invariant texture representation based on frequency decomposition and gradient orientation. Pattern Recognition Letters 51, 57–62.

Zhang, J., Marsza\lek, M., Lazebnik, S., Schmid, C., 2007. Local features and kernels for classification of texture and object categories: A comprehensive study. International journal of computer vision 73 (2), 213–238.

Zhang, Y., Tomuro, N., Furst, J., Raicu, D. S., 2010. Image enhancement and edge-based mass segmentation in mammogram. In: SPIE Medical Imaging. International Society for Optics and Photonics, pp. 76234P–76234P.

Zhang, Y., Tomuro, N., Furst, J., Raicu, D. S., 2012. Building an ensemble system for diagnosing masses in mammograms. International journal of computer assisted radiology and surgery 7 (2), 323–329.

Zhao, G., Pietikainen, M., 2007. Dynamic texture recognition using local binary patterns with an application to facial expressions. Pattern Analysis and Machine Intelligence, IEEE Transactions on 29 (6), 915–928.

Zhao, Y., Huang, D.-S., Jia, W., 2012. Completed local binary count for rotation invariant texture classification. IEEE transactions on image processing 21 (10), 4492–4497.